



(57) Abstract

(57) Zusammenfassung

Diese Erfindung betrifft regulatorische DNA-Sequenzen, beinhaltend Promotorsequenzen, sowie Intronsequenzen, für das Gen der humanen kalytischen Telomerase-Untereinheit. Darüber hinaus betrifft diese Erfindung die Verwendung dieser DNA-Sequenzen für pharmazeutische, diagnostische und therapeutische Zwecke, vor allem in der Behandlung von Krebs und Alterung.

LEDIGLICH ZUR INFORMATION

Codes zur Identifizierung von PCT-Vertragsstaaten auf den Kopfbögen der Schriften, die internationale Anmeldungen gemäss dem PCT veröffentlichen.

AL	Albanien	ES	Spanien	LS	Lesotho	SI	Slowenien
AM	Armenien	FI	Finnland	LT	Litauen	SK	Slowakei
AT	Österreich	FR	Frankreich	LU	Luxemburg	SN	Senegal
AU	Australien	GA	Gabun	LV	Lettland	SZ	Swasiland
AZ	Aserbaidschan	GB	Vereinigtes Königreich	MC	Monaco	TD	Tschad
BA	Bosnien-Herzegowina	GE	Georgien	MD	Republik Moldau	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagaskar	TJ	Tadschikistan
BE	Belgien	GN	Guinea	MK	Die ehemalige jugoslawische Republik Mazedonien	TM	Turkmenistan
BF	Burkina Faso	GR	Griechenland	ML	Mali	TR	Türkei
BG	Bulgarien	HU	Ungarn	MN	Mongolei	TT	Trinidad und Tobago
BJ	Benin	IE	Irland	MR	Mauritanien	UA	Ukraine
BR	Brasilien	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Island	MX	Mexiko	US	Vereinigte Staaten von Amerika
CA	Kanada	IT	Italien	NE	Niger	UZ	Usbekistan
CF	Zentralafrikanische Republik	JP	Japan	NL	Niederlande	VN	Vietnam
CG	Kongo	KE	Kenia	NO	Norwegen	YU	Jugoslawien
CH	Schweiz	KG	Kirgisistan	NZ	Neuseeland	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Demokratische Volksrepublik Korea	PL	Polen		
CM	Kamerun	KR	Republik Korea	PT	Portugal		
CN	China	KZ	Kasachstan	RO	Rumänien		
CU	Kuba	LC	St. Lucia	RU	Russische Föderation		
CZ	Tschechische Republik	LI	Liechtenstein	SD	Sudan		
DE	Deutschland	LK	Sri Lanka	SE	Schweden		
DK	Dänemark	LR	Liberia	SG	Singapur		
EE	Estland						

Regulatorische DNA-Sequenzen des Gens der humanen katalytischen
Telomerase-Untereinheit und deren diagnostische und therapeutische Verwen-
dung

5 Aufbau und Funktion der Chromosomenenden

Das genetische Material eukaryontischer Zellen ist auf linearen Chromosomen verteilt. Die Enden der Erbanlagen werden, abgeleitet von den griechischen Wörtern *telos* (Ende) und *meros* (Teil, Segment), als Telomere bezeichnet. Die meisten
10 Telomere bestehen aus Wiederholungen von kurzen Sequenzen, die überwiegend aus Thymin und Guanin aufgebaut sind (Zakian, 1995). In allen bislang untersuchten Wirbeltieren werden die Telomere aus der Sequenz TTAGGG aufgebaut (Meyne *et al.*, 1989).

15 Die Telomere üben verschiedene wichtige Funktionen aus. Sie verhindern die Fusion von Chromosomen (McClintock, 1941) und damit die Entstehung von dizentrischen Erbanlagen. Solche Chromosomen mit zwei Centromeren können durch Verlust der Heterozygotie bzw. Verdopplung oder Verlust von Genen zur Entwicklung von Krebs führen.

20 Desweiteren dienen Telomere dazu, intakte Erbanlagen von beschädigten zu unterscheiden. So stellten Hefezellen ihre Zellteilung ein, wenn sie ein Chromosom ohne Telomer enthielten (Sandell und Zakian, 1993).

25 Eine weitere wichtige Aufgabe erfüllen Telomere bei der DNA-Replikation eukaryontischer Zellen. Im Gegensatz zu den zirkulären Genomen von Prokaryonten können die linearen Chromosomen der Eukaryonten von dem DNA Polymerase-Komplex nicht vollständig repliziert werden. Zur Initiation der DNA-Replikation sind RNA-Primer notwendig. Nach Abspaltung der RNA-Primer, Verlängerung der
30 Okazaki-Fragmente und anschließender Ligation fehlt dem neu-synthetisierten DNA-Strang das 5'-Ende, denn dort kann der RNA-Primer nicht durch DNA ersetzt

werden. Ohne besondere Schutzmechanismen würden daher die Chromosomen mit jeder Zellteilung schrumpfen ("end-replication problem"; Harley *et al.*, 1990). Die nicht-kodierenden Telomersequenzen stellen vermutlich eine Pufferzone dar, um dem Verlust von Genen vorzubeugen (Sandell und Zakian, 1993).

5

Darüberhinaus spielen Telomere auch eine wichtige Rolle bei der Regulation der zellulären Alterung (Olovnikov, 1973). Humane somatische Zellen zeigen in Kultur eine limitierte Replikationskapazität; sie werden nach einer gewissen Zeit seneszent. In diesem Zustand teilen sich die Zellen selbst nach Stimulierung mit Wachstumsfaktoren nicht mehr, sterben aber nicht, sondern bleiben metabolisch aktiv (Goldstein, 10
1990). Verschiedene Beobachtungen sprechen für die Hypothese, daß eine Zelle anhand der Länge ihrer Telomere bestimmt, wie oft sie sich noch teilen kann (Allsopp *et al.*, 1992).

15 Zusammenfassend besitzen die Telomere somit zentrale Funktionen bei der Alterung von Zellen sowie der Stabilisierung des genetischen Materials und Verhinderung von Krebs.

Das Enzym Telomerase synthetisiert die Telomere

20

Wie oben beschrieben können Organismen mit linearen Chromosomen ohne einen speziellen Schutzmechanismus ihr Genom nur unvollständig replizieren. Die meisten Eukaryonten verwenden zur Regeneration der Telomersequenzen ein spezielles Enzym, die Telomerase. In den bislang untersuchten Einzelzellern wird Telomerase konstitutiv exprimiert. Dagegen wurde in Menschen die Telomerase-Aktivität nur in Keimzellen und Tumorzellen gemessen, wogegen benachbartes somatisches Gewebe keine Telomerase enthielt (Kim *et al.*, 1994).

25

Funktionell kann die Telomerase auch als terminale Telomertransferase bezeichnet werden, die als Multiproteinkomplex im Zellkern lokalisiert ist. Während der RNA-Anteil der humanen Telomerase schon seit längerem bekannt ist (Feng *et al.*, 1995),

30

wurde kürzlich die katalytische Untereinheit dieser Enzymgruppe in verschiedenen Organismen identifiziert (Lingner *et al.*, 1997; vgl. unsere ebenfalls anhängige Anmeldung PCT EP/98/03468). Diese katalytischen Untereinheiten der Telomerase sind sowohl untereinander als auch zu bisher allen bekannten reversen Transkriptasen auffällig homolog.

Auch in WO 98/14592 werden Nukleinsäure- und Aminosäuresequenzen der katalytischen Telomerase-Untereinheit beschrieben.

10 Aktivierung der Telomerase in menschlichen Tumoren

Eine Aktivität der Telomerase konnte in Menschen ursprünglich nur in Keimbahnzellen, nicht aber in normalen somatischen Zellen (Hastie *et al.*, 1990; Kim *et al.*, 1994) nachgewiesen werden. Nach der Entwicklung eines sensitiveren Nachweisverfahrens (Kim *et al.*, 1994) wurde auch in hematopoietischen Zellen eine geringe Telomeraseaktivität detektiert (Broccoli *et al.*, 1995; Counter *et al.*, 1995; Hiyama *et al.*, 1995). Allerdings wiesen diese Zellen trotzdem eine Reduktion der Telomere auf (Vaziri *et al.*, 1994; Counter *et al.*, 1995). Noch ist nicht geklärt, ob die Menge an Enzym in diesen Zellen nicht ausreichend für eine Kompensation des Telomerverlustes ist, oder ob die gemessene Telomerase-Aktivität von einer Subpopulation, z.B. unvollständig ausdifferenzierten CD34⁺38⁺-Vorläuferzellen, herrührt (Hiyama *et al.*, 1995). Zur Klärung wäre ein Nachweis der Telomerase-Aktivität in einer einzelnen Zelle nötig.

Interessanterweise wurde jedoch in einer großen Zahl der bislang getesteten Tumorgewebe eine signifikante Telomerase-Aktivität nachgewiesen (1734/2031, 85 %; Shay, 1997), während in normalem somatischen Gewebe keine Aktivität gefunden wurde (1/196, <1 %, Shay, 1997). Verschiedene Untersuchungen zeigten außerdem, daß in seneszenten Zellen, die mit viralen Oncoproteinen transformiert wurden, die Telomere weiterhin schrumpften und Telomerase nur in der Subpopulation entdeckt werden konnte, die die Wachstumskrise überlebte (Counter *et al.*, 1992). In diesen immortalisierten Zellen waren auch die Telomere stabil (Counter *et al.*, 1992). Ähnli-

che Befunde aus Untersuchungen an Mäusen (Blasco *et al.*, 1996) stützen die Annahme, daß eine Reaktivierung der Telomerase ein spätes Ereignis in der Tumorgenese ist.

5 Basierend auf diesen Ergebnissen wurde eine "Telomerase-Hypothese" entwickelt, die den Verlust von Telomersequenzen und Zellalterung mit der Aktivität von Telomerase und der Entstehung von Krebs verbindet. In langlebigen Spezies wie dem Menschen kann das Schrumpfen der Telomere als ein Mechanismus zur Tumorsuppression angesehen werden. Ausdifferenzierte Zellen, die keine Telomerase
10 enthalten, stellen bei einer bestimmten Länge der Telomere ihre Zellteilung ein. Mutiert eine solche Zelle, so kann aus ihr nur dann ein Tumor entstehen, wenn die Zelle ihre Telomere verlängern kann. Ansonsten würde die Zelle weiterhin Telomersequenzen verlieren, bis ihre Chromosomen instabil werden und sie schließlich zugrunde geht. Die Reaktivierung der Telomerase ist vermutlich der Hauptmechanismus
15 von Tumorzellen zur Stabilisation ihrer Telomere.

Aus diesen Beobachtungen und Überlegungen ergibt sich, daß eine Inhibition der Telomerase eine Therapie von Tumoren erlauben sollte. Konventionelle Krebstherapien mit Zytostatika oder kurzweiligen Strahlen schädigen nicht nur die Tumorzellen,
20 sondern alle sich teilenden Zellen des Körpers. Da aber außer Tumorzellen nur Keimbahnzellen eine signifikante Telomerase-Aktivität enthalten, würden Telomerase-Inhibitoren spezifischer die Tumorzellen angreifen und somit weniger unerwünschte Nebenwirkungen hervorrufen. In allen bislang getesteten Tumorgeweben wurde eine Telomerase-Aktivität nachgewiesen, so daß diese Therapeutika gegen alle
25 Krebsarten eingesetzt werden könnten. Die Wirkung von Telomerase-Inhibitoren würde dann eintreten, wenn die Telomere der Zellen sich soweit verkürzt haben, daß das Genom instabil wird. Da Tumorzellen meist kürzere Telomere aufweisen als normale somatische Zellen, würden zuerst Krebszellen durch Telomerase-Inhibitoren eliminiert werden. Zellen mit langen Telomeren, wie die Keimzellen, würden
30 dagegen erst viel später geschädigt werden. Telomerase-Inhibitoren stellen somit einen zukunftsweisenden Weg für die Therapie von Krebs dar.

Eindeutige Antworten auf die Frage nach der Art und den Angriffspunkten physiologischer Telomerase-Inhibitoren werden möglich sein, wenn auch die Regulation der Genexpression der Telomerase identifiziert ist.

5

Regulation der Genexpression in Eukaryonten

Die eukaryotische Genexpression, d.h. der zelluläre Informationsfluß von der DNA über die RNA zum Protein, weist vielfältige Ansatzpunkte für regulatorische Mechanismen auf. Einzelne Kontrollstufen sind z.B. die Gen-Amplifikation, Rekombination von Genloci, Chromatinstruktur, DNA-Methylierung, Transkription, posttranskriptionelle mRNA-Modifikationen, mRNA-Transport, Translation und posttranslationale Proteinmodifikationen. Nach bisherigen Studien besitzt die Kontrolle auf der Ebene der Transkriptionsinitiation die größte Bedeutung (Latchman, 1991).

15

Unmittelbar stromaufwärts vom Transkriptionsstart eines von der RNA-Polymerase II transkribierten Gens liegt eine Region, die für die Steuerung der Transkription verantwortlich ist und als Promotorregion bezeichnet wird. Ein Vergleich der Nukleotidsequenzen von Promotorregionen vieler bekannter Gene zeigt, daß bestimmte Sequenzmotive in dieser Region häufig vorkommen. Zu diesen Elementen gehören unter anderem die TATA-Box, die CCAAT-Box und die GC-Box, die von spezifischen Proteinen erkannt werden. Die TATA-Box, die etwa 30 Nukleotide stromaufwärts vom Transkriptionsstart entfernt positioniert ist, wird z.B. von der TFIID-Untereinheit TBP („TATA-box binding protein“) erkannt, wogegen bestimmte GC-reiche Sequenzabschnitte vom Transkriptionsfaktor Spl („specificity protein“) spezifisch gebunden werden.

25

Funktionell kann man den Promotor in einen regulativen und einen konstitutiven Abschnitt unterteilen (Latchman, 1991). Der konstitutive Kontrollbereich umfaßt den sogenannten Kernpromotor („corepromoter“), der die korrekte Initiation der Transkription ermöglicht. Er enthält die als UPE's (upstream promoter elements“) be-

30

schriebenen Sequenzelemente, die für eine effiziente Transkription notwendig sind. Die regulativen Kontrollabschnitte, die mit den UPE's verflochten sein können, weisen Sequenzelemente auf, die an der signalabhängigen Regulation der Transkription durch Hormone, Wachstumsfaktoren usw. beteiligt sein können. Sie vermitteln gewebs- oder zellspezifische Promotoreigenschaften.

Ein charakteristisches Merkmal eukaryotischer Gene sind DNA-Abschnitte, die über vergleichsweise große Distanzen hinweg Einfluß auf die Genexpression nehmen können. Diese Elemente können stromaufwärts, stromabwärts oder innerhalb einer Transkriptionseinheit lokalisiert sein und unabhängig von ihrer Orientierung ihre Funktion wahrnehmen. Diese Sequenzabschnitte können die Promotoraktivität verstärken (Enhancer) oder abschwächen (Silencer). Ähnlich wie die Promotorregionen beherbergen auch Enhancer und Silencer mehrere Bindungsstellen für Transkriptionsfaktoren.

Die Erfindung betrifft die DNA-Sequenzen aus der 5'-flankierenden Region des Gens der katalytisch aktiven humanen Telomerase-Untereinheit sowie Intron-Sequenzen für dieses Gen.

Die Erfindung betrifft insbesondere die 5'-flankierende regulatorische DNA-Sequenz, enthaltend die Promotor-DNA-Sequenz für das Gen der humanen katalytischen Telomerase Untereinheit gemäß Fig. 10 (SEQ ID NO 3).

Die Erfindung betrifft weiterhin regulatorisch wirksame Teilbereiche der 5'-flankierenden regulatorischen DNA-Sequenz gemäß Fig. 4 (SEQ ID NO 1).

Weiterhin sind Gegenstand der vorliegenden Erfindung Intron-Sequenzen für das Gen der humanen katalytischen Telomerase-Untereinheit, insbesondere solche, die regulatorische Wirkung haben. Die erfindungsgemäßen Intronsequenzen werden im Rahmen von Beispiel 5 detailliert beschrieben (vgl. SEQ ID NO 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 und 20).

Die Erfindung betrifft weiterhin ein rekombinantes Konstrukt, das die erfindungsgemäßen DNA-Sequenzen, insbesondere die 5'-flankierende DNA-Sequenz des Gens der humanen katalytischen Telomerase Untereinheit oder Teilbereiche davon
5 beinhaltet.

Bevorzugt sind rekombinante Konstrukte, die neben den erfindungsgemäßen DNA-Sequenzen, insbesondere der 5'-flankierenden DNA-Sequenz des Gens der humanen katalytischen Telomerase Untereinheit oder Teilbereichen davon, eine oder mehrere
10 weitere DNA-Sequenzen, die für Polypeptide oder Proteine kodieren, enthalten.

Gemäß einer besonders bevorzugten Ausführungsform kodieren diese weiteren DNA-Sequenzen für antitumorale Proteine.

15 Besonders bevorzugte antitumorale Proteine sind solche, die die Angiogenese direkt oder indirekt inhibieren. Zu diesen Proteinen zählen beispielsweise:

Plasminogenaktivatorinhibitor (PAI-1), PAI-2, PAI-3, Angiostatin, Endostatin, Platelet factor 4, TIMP-1, TIMP-2, TIMP-3, Leukemia Inhibitory Factor (LIF).
20

Ebenfalls besonders bevorzugt sind antitumorale Proteine, welche direkt oder indirekt eine zytostatische Wirkung auf Tumoren aufweisen. Hierzu zählen im besonderen:

25 Perforin, Granzym, IL-2, IL-4, IL-12, Interferone, wie beispielsweise IFN- α , IFN- β , IFN- γ , TNF, TNF- α , TNF- β , Oncostatin M; Tumorsuppressorgene, wie z.B. p53, Retinoblastoma.

Weiterhin besonders bevorzugt sind antitumorale Proteine, welche gegebenenfalls
30 zusätzlich zur antitumoralen Wirkung Entzündungen stimulieren und hierdurch zur Elimination von Tumorzellen beitragen. Hierzu zählen beispielsweise:

RANTES, Monocyte chemotactic and activating factor (MCAF), IL-8, Macrophage inflammatory protein (MIP-1 α , - β), Neutrophil activating protein-2 (NAP-2), IL-3, IL-5, human leukemia inhibitory factor (LIF), IL-7, IL-11, IL-13, GM-CSF, G-CSF, M-CSF.

Weiterhin besonders bevorzugt sind antitumorale Proteine, welche aufgrund ihrer Wirkung als Enzyme in der Lage sind, Vorstufen eines antitumoralen Wirkstoffes in einen antitumoralen Wirkstoff zu überführen. Zu diesen Enzymen zählen beispielsweise:

Herpes Simplex Virus Thymidinkinase, Varizella Zoster Virus Thymidinkinase, bakterielle Nitroreductase, bakterielle β -Glukuronidase, pflanzliche β -Glukuronidase aus *Secale cereale*, humane Glukuronidase, humane Carboxypeptidase, bakterielle Carboxypeptidase, bakterielle β -Lactamase, bakterielle Cytosindeaminidase, humane Katalase bzw. Phosphatase, humane alkalische Phosphatase, Typ 5 saure Phosphatase, humane Lysooxidase, humane saure D-Aminooxidase, humane Glutathion Peroxidase, humane Eosinophilen Peroxidase, humane Schilddrüsen Peroxidase.

Die obengenannten rekombinanten Konstrukte können auch DNA-Sequenzen enthalten, die für Faktor VIII, IX oder Teilfragmente davon kodieren. Zu diesen DNA-Sequenzen zählen auch andere Blutgerinnungsfaktoren

Die obengenannten rekombinanten Konstrukte können auch DNA-Sequenzen enthalten, die für ein Reporterprotein kodieren. Zu diesen Reporterproteinen zählen beispielsweise:

Chloramphenicolacetyltransferase (CAT), Glühwürmchen Luziferase (LUC), β -Galaktosidase (β -Gal), Sezernierte alkalische Phosphatase (SEAP), Humanes Wachstumshormon (hGH), β -Glukuronidase (GUS), Grün-fluoreszierendes Protein (GFP) und alle davon abgeleiteten Varianten, Aquarin, Obelin.

Erfindungsgemäße rekombinante Konstrukte können auch DNA kodierend für die humane katalytische Telomerase Untereinheit und deren Varianten und Fragmente in antisense Orientierung enthalten. Gegebenenfalls können diese Konstrukte auch
5 andere Protein-Untereinheiten der humanen Telomerase und die Telomerase-RNA-Komponente in antisense Orientierung enthalten.

Die rekombinanten Konstrukte können neben der DNA, kodierend für die humane katalytische Telomerase Untereinheit, sowie deren Varianten und Fragmente auch
10 andere Protein-Untereinheiten der humanen Telomerase und die Telomerase-RNA-Komponente enthalten.

Die Erfindung betrifft weiterhin einen Vektor, enthaltend die oben genannten erfindungsgemäßen DNA-Sequenzen, insbesondere die 5'-flankierenden DNA-Sequenzen,
15 zen, sowie eine oder mehrere der oben genannten anderen DNA-Sequenzen.

Bevorzugter Vektor für solche Konstrukte ist ein Virus, beispielsweise ein Retrovirus, Adenovirus, adeno-assoziiertes Virus, Herpes Simplex Virus, Vaccina Virus, lentivirales Virus, Sindbis Virus und ein Semliki Forest Virus.
20

Ebenfalls bevorzugt sind Plasmide als Vektoren.

Die Erfindung betrifft weiterhin pharmazeutische Präparate, enthaltend erfindungsgemäße rekombinante Konstrukte bzw. Vektoren; beispielsweise eine Zubereitung in einem kolloidalen Dispersionssystem.
25

Geeignete kolloidale Dispersionssysteme sind beispielsweise Liposome oder Polylysin-Liganden.

30 Die Zubereitungen der erfindungsgemäßen Konstrukte bzw. Vektoren in kolloidalen Dispersionssystemen können um einen Liganden ergänzt sein, der an Membranstruk-

turen von Tumorzellen bindet. Ein solcher Ligand kann z.B. an das Konstrukt bzw. den Vektor angeknüpft sein oder auch Bestandteil der Liposomenstruktur sein.

5 Geeignete Liganden sind insbesondere polyklonale oder monoklonale Antikörper oder Antikörperfragmente hiervon, die mit ihren variablen Domänen an Membranstrukturen von Tumorzellen binden, oder endständige Mannose-tragende Substanzen, Zytokine, Wachstumsfaktoren oder Fragmente bzw. Teilsequenzen hiervon, die an Rezeptoren auf Tumorzellen binden.

10 Entsprechende Membranstrukturen sind beispielsweise Rezeptoren für ein Zytokin oder einen Wachstumsfaktor, wie z.B. IL-1, EGF, PDGF, VEGF, TGF β , Insulin oder Insulin-like Growth Factor (ILGF), oder Adhäsionsmoleküle, wie z. B. SLeX, LFA-1, MAC-1, LECAM-1 oder VLA-4, oder der Mannose-6-Phosphat-Rezeptor.

15 Zur vorliegenden Erfindung gehören pharmazeutische Zubereitungen, die neben den erfindungsgemäßen Vektorkonstrukten auch nichttoxische, inerte, pharmazeutisch geeignete Trägerstoffe enthalten können. Vorstellbar sind die Applikation (z.B. intravenös, intraarteriell, intramuskulär, subkutan, intradermal, anal, vaginal, nasal, transdermal, intraperitoneal, als Aerosol oder oral) am Ort eines Tumors oder die systemische Applikation dieser Zubereitungen.

20

Die erfindungsgemäßen Vektorkonstrukte können in der Gentherapie eingesetzt werden.

25 Die Erfindung betrifft weiterhin eine rekombinante Wirtszelle, insbesondere eine rekombinante eukaryotische Wirtszelle, enthaltend die vorstehend beschriebenen Konstrukte bzw. Vektoren.

30 Die Erfindung betrifft weiterhin ein Verfahren zur Identifizierung von Substanzen, die die Promotor-, Silencer- oder Enhanceraktivität der katalytischen Telomerase Untereinheit beeinflussen, wobei dieses Verfahren folgende Schritte umfaßt:

A. Zugabe einer Kandidatensubstanz zu einer Wirtszelle, enthaltend die erfindungsgemäße regulatorische DNA-Sequenz, insbesondere die 5'-flankierende regulatorische DNA-Sequenz für das Gen der humanen katalytischen
5 Telomerase-Untereinheit oder einen regulatorisch wirksamen Teilbereich davon, funktionell verknüpft mit einem Reportergen,

B. Messung des Substanzeffektes auf die Reportergenexpression.

10 Das Verfahren kann eingesetzt werden zur Identifizierung von Substanzen, die die Promotor-, Silencer- oder Enhanceraktivität der katalytischen Telomerase Untereinheit verstärken.

Das Verfahren kann weiterhin eingesetzt werden zur Identifizierung von Substanzen,
15 die die Promotor-, Silencer- oder Enhanceraktivität der katalytischen Telomerase Untereinheit inhibieren.

Die Erfindung betrifft weiterhin ein Verfahren zur Identifizierung von Faktoren, die spezifisch an Fragmente der erfindungsgemäßen DNA-Fragmente, insbesondere der
20 5'-flankierenden regulatorischen DNA-Sequenz der katalytischen Telomerase Untereinheit, binden. Diese Methode beinhaltet ein Screening einer Expressions-cDNA-Bibliothek mit der vorstehend beschriebenen DNA-Sequenz oder Teilfragmenten unterschiedlichster Länge als Sonde.

25 Die vorstehend beschriebenen Konstrukte bzw. Vektoren können auch zur Herstellung transgener Tiere verwendet werden.

Die Erfindung betrifft weiterhin ein Verfahren zur Detektion Telomerase-assoziiierter Zustände bei einem Patienten, das folgende Schritte umfaßt:

30

- A. Inkubation eines Konstruktes bzw. Vektors, enthaltend die erfindungsgemäße DNA-Sequenz, insbesondere die 5'-flankierende regulatorische DNA-Sequenz für das Gen der humanen katalytischen Telomerase-Untereinheit oder einen regulatorisch wirksamen Teilbereich davon sowie ein Reporterogen mit Körperflüssigkeiten oder zellulären Proben,
- B. Detektion der Reporterogenaktivität, um einen diagnostischen Wert zu erhalten;
- C. Vergleich des diagnostischen Werts mit Standardwerten für das Reportergenkonstrukt in standardisierten normalen Zellen oder Körperflüssigkeiten des gleichen Typs wie die Testprobe;

Detektion diagnostischer Werte, die höher oder niedriger als Standardvergleichswerte liegen, indiziert einen Telomerase-assoziierten Zustand, der wiederum einen pathogenen Zustand indiziert.

Erläuterung der Abbildungen:

- Fig. 1: Southern Blot-Analyse mit genomischer DNA verschiedener Spezies

A: Foto eines Ethidiumbromid gefärbten 0,7 %igen Agarosegels mit etwa 4 µg Eco RI geschnittener genomischer DNA. Die Spur 1 enthält Hind III geschnittene λ-DNA als Größenmarker (23,5, 9,4, 6,7, 4,4, 2,3, 2,0, und 0,6 kb). Die Spuren 2 bis 10 enthalten genomische DNA von Mensch, Rhesusaffe, Sprague Dawley Ratte, BALB/c Maus, Hund, Rind, Kaninchen, Huhn und Hefe (*Saccharomyces cerevisiae*).

B: Zu Fig.1 A korrespondierendes Autoradiogramm einer Southern Blot-Analyse, hybridisiert mit einer radioaktiv-markierten etwa 720 bp langen hTC-cDNA Sonde.

Fig. 2: Restriktionsanalyse der rekombinanten λ -DNA des Phagenklons P12, der mit einer Sonde aus dem 5'-Bereich der hTC-cDNA hybridisiert.

5 Die Abbildung zeigt ein Foto eines Ethidiumbromid gefärbten 0,4 %igen Agarosegels. Die Spuren 1 und 2 enthalten Eco RI/Hind III geschnittene λ -DNA bzw eine 1 kb Leiter der Firma Gibco als Größenmarker. Die Spuren 3 - 7 enthalten 250 ng mit Bam HI (Spur 3), Eco RI (Spur 4), Sal I (Spur 5), Xho I (Spur 6) und Sac I (Spur 7) geschnittene DNA des
10 rekombinanten Phagens. Die Pfeile kennzeichnen die zwei λ -Arme des Vektors EMBL3 Sp6/T7.

Fig. 3: Restriktionsanalyse und Southern Blot-Analyse der rekombinanten λ -DNA des Phagenklons, der mit einer Sonde aus dem 5'-Bereich der hTC-cDNA hybridisiert.

15 A: Die Abbildung zeigt ein Foto eines Ethidiumbromid gefärbten 0,8%igen Agarosegels. Die Spuren 1 und 15 enthalten eine 1 kb Leiter der Firma Gibco als Größenmarker. Die Spuren 2 bis 14 enthalten 250 ng geschnittene λ -DNA vom rekombinanten Phagenklon. Als Enzyme wurden eingesetzt: Spur 2: Sac I, Spur 3: Xho I, Spur 4: Xho I, Xba I, Spur 5: Sac I, Xho I, Spur 6: Sal I, Xho I, Xba I, Spur 7: Sac I, Xho I, Xba I, Spur 8: Sac I, Sal I, Xba I, Spur 9: Sac I, Sal I, BamH I, Spur 10: Sac I, Sal I, Xho I, Spur 11: Not I, Spur 12: Sma I, Spur 13: leer, Spur
20 14: nicht verdaut.

25 B: Zu Fig.3 A korrespondierendes Autoradiogramm einer Southern Blot-Analyse. Als Sonde für die Hybridisierung wurde ein etwa 420 bp langes 5'-hTC-cDNA Fragment eingesetzt.

30

Fig. 4: Partielle DNA-Sequenz der 5'-flankierenden Region und des Promotors vom Gen der humanen katalytischen Telomerase-Untereinheit. Das ATG-Startcodon ist in der Sequenz fett hervorgehoben. Die dargestellte Sequenz entspricht SEQ ID NO 1.

5

Fig. 5: Identifizierung des Transkriptionsstarts durch Primer Extension-Analyse.

10

Die Abbildung zeigt ein Autoradiogramm eines denaturierenden Polyacrylamidgels, welches zur Darstellung einer Primer Extension-Analyse gewählt wurde. Als Primer wurde ein Oligonukleotid mit der Sequenz 5'GTTAAGTTGTAGCTTACACTGGTTCTC 3' benutzt. In der Spur 1 wurde die Primer Extension Reaktion aufgetragen. Die Spuren G, A, T, C, stellen die Sequenzreaktionen mit dem gleichen Primer und den entsprechenden Dideoxynukleotiden dar. Der fette Pfeil kennzeichnet den Haupt-Transkriptionsstart, die dünnen Pfeile weisen auf drei Neben-Transkriptionsstartpunkte hin.

15

Fig. 6: cDNA Sequenz der humanen katalytischen Telomerase-Untereinheit (hTC; vgl. unsere anhängige Anmeldung PCT/EP/98/03468). Die dargestellte Sequenz entspricht SEQ ID NO 2.

20

Fig. 7: Strukturelle Organisation und Restriktionsmappe des humanen hTC-Gens und dessen 5'- und 3'-flankierende Region.

25

Exons sind als durchnummerierte schwarz ausgefüllte Rechtecke und Introns als nicht ausgefüllte Bereiche hervorgehoben. Nichttranslatierte Sequenzabschnitte in den Exons sind schraffiert. Die Translation startet in Exon 1 und endet in Exon 16. Restriktionsenzymchnittstellen sind wie folgt gekennzeichnet: S, SacI; X, XhoI. Die relative Anordnung der fünf Phagenklone (P2, P3, P5, P12, P17) und des Produktes aus dem „Genomic walking“ sind durch dünne Linien hervorgehoben. Wie durch

30

die Punkte gekennzeichnet, ist die Sequenz von Intron 16 nur teilweise entschlüsselt.

Fig. 8: HTL Splicevarianten.

5

10

15

A: Schematische Struktur der hTC mRNA Splicevarianten. Die vollständige hTC mRNA ist als grau unterlegtes Rechteck im oberen Bereich der Abb. dargestellt. Die 16 Exons sind entsprechend ihrer Größe dargestellt. Der Translationsstart (ATG) und das Stop-Codon, sowie das Telomerase-spezifische T-Motiv und die sieben RT-Motive sind hervorgehoben. Die hTC-Varianten sind in Deletions- und Insertionsvarianten unterteilt. In den Deletionen sind die fehlenden Exonsequenzen markiert. Die Insertionen sind durch zusätzliche weiße Rechtecke hervorgehoben. Größe und Herkunft der insertierten Sequenzen sind angegeben. Neu entstandene Stop-Codons sind markiert. Die Größe der Insertion von Variante INS2 ist unbekannt.

20

25

B: Exon Intron Übergänge der hTC-Splice-Varianten. Nichtgespligte 5'- und 3'-flankierte Sequenzen sind als weiße Rechtecke hervorgehoben. Die Herkunft der Exon und Intron Sequenzen ist angegeben. Intron und Exon Sequenzen sind in Kleinbuchstaben, bzw. in Großbuchstaben dargestellt. Die Donor und Akzeptor Sequenzen der Splicestellen sind als graue Rechtecke unterlegt und deren Exon Intron Herkunft ist ebenfalls angegeben.

Fig. 9: Identifizierung des Transkriptionsstarts durch RT-PCR Analyse.

30

Die RT-PCR wurde mit cDNA-Bibliothek aus HL 60 Zellen und genomischer DNA als Positivkontrolle durchgeführt. Ein gemeinsamer 3'-Primer hybridisiert an eine Sequenzregion aus Exon 1. Die Position der verschiedenen 5' Primer in der kodierenden Region oder der 5'-flankierenden Region ist angegeben. In der Negativkontrolle wurde keine

Template-DNA in der PCR-Reaktion zugegeben. M: DNA-Größenmarker.

Fig. 10: Nukleotidsequenz und Strukturmerkmale des hTC-Promotors.

5 11273 bp der 5'-flankierenden hTC Gensequenz, beginnend mit dem Translationsstartcodon ATG (+1) sind dargestellt. Die putative Region des Translationsstarts ist unterstrichen. Mögliche regulatorische Sequenzabschnitte innerhalb der 4000 bp stromaufwärts des Translationsstarts sind umrandet. Die dargestellte Sequenz entspricht SEQ ID NO 3.

10

Fig. 11: Aktivität des hTC-Promotors in HEK-293 Zellen.

Im oberen Bereich der Abbildung sind die ersten 5000 bp der 5'-flankierenden hTC Genregion schematisch dargestellt. Das ATG-Startcodon ist hervorgehoben. CpG reiche Inseln sind durch graue Rechtecke markiert. Auf der linken Abbildungsseite sind die Größen der hTC Promotor-Luziferase Konstrukte dargestellt. Das promotorlose pGL2-Basic Konstrukt und das SV40 Promotorkonstrukt pGL2-Pro wurden in jeder Transfektion als Kontrollen eingesetzt. Auf der rechten Abbildungsseite ist die relative Luziferaseaktivität der verschiedenen Promotorkonstrukte in HEK-Zellen als durchgehende Balken gezeigt. Die Standardabweichung ist angegeben. Die Zahlenwerte repräsentieren den Durchschnitt von zwei unabhängigen Experimenten, die in Duplikaten durchgeführt wurden.

25 Tab. 1: Exon Intron Übergänge des hTC-Gens

Aufgelistet sind die Nukleotidsequenzen an den 3'- und 5' Spliceübergängen des hTC-Gens. Die Konsensussequenzen für Donor und Akzeptorsequenzen (AG und GT) sind durch graue Rechtecke unterlegt. Intronsequenzen (Kleinbuchstaben) und Exonsequenzen (Großbuchstaben), die die Spliceakzeptor- und Donorstellen flankieren sind
30 gezeigt. Die Größe der Exons und Introns ist in bp angegeben.

Tab. 2: Potentielle Bindungsstellen für DNA-bindende Faktoren in der Nukleotidsequenz von Intron 2

5 Die Suche nach möglichen DNA-bindenden Faktoren (z.B. Transkriptionsfaktoren) wurde mit dem „Find Pattern“-Algorithmus aus dem „GCG Sequenz Analysis“ Programmpaket der „Genetics Computer Group“ (Madison, USA) durchgeführt. Aufgelistet sind die Abkürzungen der identifizierten DNA-bindenden Faktoren und deren Lokalisation in

10 Intron 2.

Tab. 1

3' Acceptor Sequence			5' Donor Sequence					
Intron	Exon	Exon No.	bp	Exon	Intron	Intron on	bp	
						No.		No.
5' flanking Region								
caggggcctcccccgtag	GTTCAGGCAGCGCTGCGT	1	281	CGCCCCCTCCTTCGCCAG	gtgggcctccccgggtcg	1	104	
catgtcctctctgttttaag	GTGTCTGCTGCTGAAGGAC	2	1354	TGGCTGCGCAGGAGCCAG	gtgaggaggtggtggccgt	2	8616	
gaggggtctctctattgtag	GGGTGGCTGTGTTCGGC	3	196	TGCAAGCATTTGGAATCAG	gtactgtatccccacgcca	3	2089	
cccatgtgtccccgcgtag	ACAGCACTTGAAGAGGCTG	4	181	GTTCCGCGAGAGAAAAGAGG	gtgctgtgcttttggttta	4	687	
ctcgccctccactcacadag	GCCGAGGCTCTCACCTCGA	5	180	TGAGCTGTACTTTTGTCAAG	gtgggtgcgggggaccccc	5	494	
ccctctcctctgcggtag	GTGDAITGTGACGGGCGCGT	6	156	CAAGGCCTTCAAGAGCCAC	gtagggttcacgtgtgata	6	>4660	18
ctccccgtctgttttcgtag	GTCTCTACCTTGACAGACC	7	96	TGCCGTGTCATCGAGCAG	gtctgggacactgccccgca	7	980	
ctggtgtcttcccccgtag	AGCTCTCCCTGATGAGG	8	86	CCGTGCGCATCAGGGGCAA	gtgagtcagggtggccagggt	8	2484	
gtattttcccttatttttag	GTCTACGTCCAGTGCCAG	9	114	CGGGGATTGCGCGGGACGG	gtgaggcctcctcttcccc	9	1984	
cattgccccctctgcttag	GCTGCTCCTGCGTTTGGTG	10	72	ACGGGAAACCTTCCTCAG	gtgagggccgtgcccgtgtg	10	1871	
attccccctgtgtcttag	GACCTGGTCCGAGGTGC	11	189	TGCAGAGCGACTACTCCAG	gtgagcgacactggcccgga	11	3804	
tcctttctggcgactcttag	CTATGCCGGACCTCCATC	12	127	CCTGTTTCTGGATTTCAG	gtgagcaggctgatggtca	12	880	
ctgtccgcacatcctcttag	GTGAACAGCCCTCCAGACGG	13	62	TCCTGCTGCAGGCGTACAG	gtgagccgccaccacaagggg	13	3184	
agcctctgttttcccccttag	GTTCACGCATGTGTGCTG	14	125	CTGAAAGCCAGAACGCCAG	gtatgtcagggtgcctgtgc	14	781	
tcgtgattttggcccccgtag	GGATGCTGCTGGGGCCAA	15	138	CTGGGCTCACCTCAGGACAG	gcaagtgtgggtggaggcc	15	536	
	CCCAGACGCAGCTGAGTCG	16	664	TTTTTCAGTTTGGAAAAA				3' flanking Region

Tab. 2

Faktoren	Lokalisation in Intron 2
C/EBP	2925
CRE.2	2749
Sp1	2378, 4094, 4526, 4787, 4835, 4995
AP-2 CS3	5099
AP-2 CS4	2213, 3699, 4667, 5878, 5938, 6059, 6180, 6496
AP-2 CS5	5350, 5798, 5880, 5940, 6061, 6182, 6375, 6498
PEA3	934, 2505
P53	2125
GR uteroglobin	848, 1487, 2956
PR uteroglobin	3331
Zeste-white	1577, 1619, 1703, 1745, 1787, 1829, 1871, 1913, 1955, 1997, 2039, 2081, 3518, 3709, 4765, 5014, 5055
GRE	846
MyoD-MCK right site/rev	447, 509, 558, 1370, 1595, 1900, 2028, 2099, 4557
MyoD-MCK left site	108, 118, 453, 1566, 1608, 1692, 1734, 1818, 1902, 1986, 2372, 2460, 2720, 3491, 5030
Ets-1 CS	6408
API	3784, 4406
CREB	2801
GATA-1	839, 1390, 3154
c-Myc	108, 118, 453, 1566, 1608, 1692, 1734, 1818, 1902, 1986, 2372, 2460, 2720, 3491, 5030
CACCC site	991
CCAAT site	1224
CCAC box	992
CAAT site	463, 2395
Rb site	992, 4663
TATA	3650
CDEI	106, 1564, 1606, 1690, 1732, 1816, 1900, 1984

Beispiele

Das menschliche Gen für die katalytische Telomerase Untereinheit (ghTC), sowie die 5' und 3' liegenden Bereiche dieses Gens wurden kloniert, der Startpunkt der Transkription bestimmt, potentielle Bindungsstellen für DNA-bindende Proteine identifiziert, sowie aktive Promotorfragmente aufgezeigt. Die Sequenz der hTC-cDNA (Fig. 6) ist bereits in unserer ebenfalls anhängigen Anmeldung PCT/EP/98/03468 beschrieben. Wenn nicht gesondert erwähnt, beziehen sich sämtliche Angaben zur cDNA-Position auf diese Sequenz.

Beispiel 1

Durch eine genomische Southern Blot-Analyse wurde bestimmt, ob ghTC im menschlichen Genom ein Einzelgen darstellt oder mehrere Loci für das hTC-Gen bzw. eventuell auch ghTC-Pseudogene existieren.

Hierzu wurde ein kommerziell erhältlicher Zoo-Blot der Firma Clontech einer Southern Blot-Analyse unterzogen. Dieser Blot enthält 4 µg Eco RI geschnittene genomische DNA von neun verschiedenen Spezies (Mensch, Affe, Ratte, Maus, Hund, Rind, Kaninchen, Huhn und Hefe). Mit Ausnahme von Hefe, Huhn und Mensch wurde die DNA aus Nierengewebe isoliert. Die humane genomische DNA wurde aus Plazenta isoliert und die genomische DNA aus Huhn wurde aus Lebergewebe aufgereinigt. Im Autoradiogramm in Fig. 1 wurde als radioaktiv-markierte Sonde ein etwa 720 bp langes hTC-cDNA Fragment, isoliert aus der hTC cDNA, Variante Del2 (Position 1685 bis 2349 plus 2531 bis 2590 der Fig. 6 [Deletion 2; vergl. Beispiel 5 der Fig. 8]), eingesetzt. Die experimentellen Bedingungen für die Hybridisierung und die Waschschrte des Blots erfolgten in Anlehnung an Ausubel *et al.* (1987).

Im Fall der humanen DNA erkennt die Sonde zwei spezifische DNA-Fragmente. Das kleinere, etwa 1,5 bis 1,8 kb lange Eco RI-Fragment geht wahrscheinlich auf zwei

Eco RI-Schnittstellen in einem Intron der ghTC-DNA zurück. Aufgrund dieses Ergebnisses ist davon auszugehen, daß nur ein singuläres ghTC-Gen im menschlichen Genom vorliegt.

5 Beispiel 2

Zur Isolierung der 5' flankierenden hTC-Gensequenz wurden ca $1,5 \times 10^6$ Phagen einer humanen genomischen Plazenta-Genbibliothek (EMBL 3 SP6/T7 der Firma Clontech, Bestellnummer HL1067j) auf Nitrozellulosefilter (0,45 µm; Fa. Schleicher und Schuell) nach Angaben des Herstellers mit einem radioaktiv markierten, etwa
10 500 bp langen 5'-hTC-cDNA Fragment (Position 839 bis 1345 der Fig. 6) hybridisiert. Die Nitrozellulosefilter wurden zunächst in 2 x SSC (0,3 M NaCl; 0,5 M Tris-HCl, pH 8,0) und anschließend in einer Prähybridisierungslösung (50 % Formamid; 5 x SSPE, pH 7,4; 5 x Denhards-Lösung; 0,25 % SDS; 100 µg/ml
15 Heringssperma-DNA) zwei Stunden bei 42°C inkubiert. Für die Hybridisierung über Nacht wurde die Prähybridisierungslösung mit $1,5 \times 10^6$ cpm/ml Lösung denaturierter, radioaktiv markierter Probe ergänzt. Unspezifisch gebundene, radioaktive DNA wurde unter stringenten Bedingungen, d.h. durch drei fünfminütige Waschschritte mit 2 x SSC; 0,1 % SDS bei 55 bis 65 °C entfernt. Die Auswertung erfolgte
20 durch Autoradiographie der Filter.

Die in dieser Primäruntersuchung identifizierten Phagenklone wurden aufgereinigt Ausubel *et al.* (1987). In weitergehenden Analysen stellte sich ein Phagenklon P12 als potentiell positiv heraus. Eine λ-DNA Präparation dieses Phagens Ausubel *et al.*
25 (1987) und der nachfolgende Restriktionsverdau mit Enzymen, die das genomische Insert in Fragmenten freisetzen, zeigte, daß dieser Phagenklon ein ca. 15 kb Insert im Vektor enthält (Fig. 2).

Zur Isolierung der vollständigen hTC-Gensequenz wurden in unabhängigen
30 Experimenten jeweils 1 bis $1,5 \times 10^6$ Phagen mit jeweils verschiedenen radioaktiv markierten Sonden wie oben beschrieben durchmustert.

Die in diesen Primäruntersuchungen identifizierten, für die entsprechenden Sonden positiven Phagenklone wurden aufgereinigt. Der Phagenklon P17 wurde mit einem etwa 250 bp langen hTC-cDNA Fragment (Position 1787 bis 2040 der Fig. 6) gefunden. Der Phagenklon P2 wurde mit einem etwa 740 bp langen hTC-cDNA Fragment (Position 1685 bis 2349 plus 2531 bis 2607 der Fig. 6 [Deletion 2; vergl. Beispiel 5]) identifiziert. Die Phagenklone P3 und P5 wurden mit einem 420 bp langen 3' hTC-cDNA Fragment (Position 3047 bis 3470 der Fig. 6) gefunden. Nach λ -DNA Präparation dieser Phagen und nachfolgendem Restriktionsverdau mit Enzymen, die das genomische Insert in Fragmenten freisetzen, wurden die Inserts in Plasmide umklontiert (Beispiel 4).

Beispiel 3

Um zu untersuchen, ob auch das 5'-Ende der hTC-cDNA im Insert des rekombinanten Phagenklons P12 vorliegt, wurde λ -DNA dieses Klons in einer Southern Blot Analyse mit einem radioaktiv markierten etwa 440 bp langen hTC-cDNA Fragment (Position 1 bis 440 der Fig. 6) aus dem extremen 5'-Bereich hybridisiert (Fig. 3).

Da die isolierte λ -DNA des positiven Klons auch mit dem extremen 5'-Ende der hTC-cDNA hybridisiert, enthält dieser Phage wahrscheinlich auch den das ATG-Startcodon flankierenden 5'-Sequenzbereich.

Beispiel 4

Um das gesamte 15 kb lange Insert des positiven Phagenklons P12 in Teilfragmenten umzuklonieren und anschließend zu sequenzieren, wurden zum DNA-Verdau Restriktionsendonukleasen ausgewählt, die zum einem das gesamte Insert aus EMBL3 Sp6/T7 freisetzen (vgl. Beispiel 2) und zusätzlich im Insert schneiden.

- Insgesamt wurden ein etwa 8,3 und ein etwa 6,5 kb langes Xho I-Subfragment sowie ein etwa 8,5, ein etwa 3,5 und ein etwa 3 kb langes Sac I-Teilfragment in den Vektor pBluescript KS(+) (Fa. Stratagene) umklont. Durch Sequenzanalyse dieser Fragmente wurde die Nukleotidsequenz von 5123 bp 5'-flankierenden des ghTC-Genbereichs, ausgehend vom ATG-Startcodon bestimmt (Fig. 4; entsprechend SEQ ID NO 1). In der Fig. 4 sind die ersten (ausgehend vom ATG-Startcodon) 5123 bp dargestellt. In der Fig. 10 (entsprechend SEQ ID NO 3) die gesamte klonierte 5' Sequenz.
- Um das gesamte ca. 14,6 kb große Insert des Phagenklons P17 in Teilfragmenten umzuklonieren, wurden zum DNA-Verdau Restriktionsendonukleasen ausgewählt, die zum einen das gesamte Insert aus EMLB3 Sp6/T7 freisetzen und zusätzlich einige Male im Insert schneiden. Durch Kombinationsverdau mit den Enzymen XhoI und BamHI wurden ein 7,1 kb, ein 4,2 kb und ein 1,5 kb großes XhoI-BamHI-Fragment sowie ein 1,8 kb großes BamHI-Fragment subklont. Der Kombinations-Restriktionsverdau mit den Enzymen XhoI und XbaI führte zur Klonierung von einem 6,5 kb großen XhoI-XbaI-Fragment, einem 6,5 kb und einem 1,5 kb großem XhoI-Fragment.
- Die Umklonierung des ca. 17,9 kb großem Inserts des Phagenklons P2 in Subfragmente erfolgte durch Verdau mit dem Restriktionsenzym XhoI. Insgesamt wurde ein 7,5 kb, ein 6,4 kb sowie ein 1,6 kb langes XhoI-Subfragment kloniert. Durch Verdau mit dem Restriktionsenzym SacI wurde zusätzlich ein 4,8 kb, ein 3 kb, ein 2 kb sowie ein 1,8 kb großes SacI-Fragment subklont.
- Das ca. 13,5 kb große Insert des Phagenklons P3 wurde durch Verdau mit den Restriktionsenzymen SacI bzw. XhoI subklont. Dabei wurden ein 3,2 kb, ein 2 kb, ein 0,9 kb, ein 0,8 kb, ein 0,65 kb und ein 0,5 kb langes SacI-Subfragment sowie ein 6,5 kb und ein 4,3 kb langes XhoI-Subfragment erhalten.

Die Subklonierung des ca. 13,2 kb großen Inserts des Phagenklons P5 erfolgte durch Verdau mit den Restriktionsenzymen SacI bzw. XhoI. Insgesamt wurden SacI-Fragmente von 6,5 kb, 3,3 kb, 3,2 kb, 0,8 kb und 0,3 kb Größe sowie XhoI-Fragmente von 7 kb und 3,2 kb Größe subkloniert.

5

Zur Klonierung des 3' von Phagenklon P17 und 5' von Phagenklon P2 gelegenen hTC-genomischen Sequenzbereichs wurden 3 Genomic Walkings mit Hilfe des GenomeWalker™ Kits der Firma Clontech (Katalognummer K1803-1) und verschiedenen Primerkombinationen durchgeführt. In einem Endvolumen von 50 µl wurde 1 µl humaner GenomeWalker Library HDL (Fa. Clontech) mit 10 pmol

10

dNTP-Mix versetzt und in 1xKlen Taq PCR-Reaktionspuffer und 1xAdvantage Klen Taq Polymerase Mix (Fa. Clontech) eine PCR-Reaktion durchgeführt. Als Primer wurden 10 pmol eines internen genspezifischen Primers sowie 10 pmol des Adaptor Primers AP1 (5'-GTAATACGACTCACTATAGGGC-3'; Fa. Clontech) zugefügt.

15

Die PCR wurde als Touchdown-PCR in 3 Schritten durchgeführt. Zunächst wurde über 7 Zyklen für 20 sec bei 94°C denaturiert und anschließend für 4 min bei 72°C die Primer angelagert und die DNA-Kette verlängert. Es folgten 37 Zyklen bei denen für 20 sec die DNA bei 94°C denaturiert wurde, die anschließende Primerverlängerung aber für 4 min bei 67°C erfolgte. Abschließend folgte eine Kettenverlängerung für 4 min bei 67°C. Im Anschluß an diese erste PCR wurde das PCR-

20

Produkt 1:50 verdünnt. Ein µl dieser Verdünnung wurde in einer zweiten „nested“ PCR zusammen mit 10 pmol dNTP-Mix in 1xKlen Taq PCR-Reaktionspuffer und 1xAdvantage Klen Taq Polymerase-Mix sowie 10 pmol eines „nested“ genspezifischen Primers und 10 pmol des „nested“ Marathon Adaptor Primers AP2 (5'-

25

ACTATAGGGCACGCGTGGT-3'; Fa. Clontech) eingesetzt. Die PCR-Bedingungen entsprachen den in der ersten PCR gewählten Parametern. Als einzige Ausnahme wurden im ersten PCR-Schritt statt 7 Zyklen nur 5 Zyklen gewählt und im zweiten PCR-Schritt statt 37 Zyklen nur 24 Zyklen durchlaufen. Produkte dieser Nested-GenomicWalking-PCR wurden in den TA-Cloning Vektor pCRII der Fa. InVitrogen

30

kloniert.

Im ersten Genomic Walking wurde der genspezifische Primer C3K2-GSP1 (5'-GACGTGGCTCTTGAAGGCCTTG-3') sowie der „nested“ genspezifische Primer C3K2-GSP2 (5'-GCCTTCTGGACCACGGCATAACC-3') zusammen mit der HDL-Library 4 eingesetzt und ein 1639 bp langes PCR-Fragment erhalten. Im zweiten
5 Genomic Walking wurde mit dem genspezifischen Primer C3F2 (5'-CGTAGTTGAGCACGCTGAACAGTG-3') und dem „nested“ genspezifischen Primer C3F (5'-CCTTCACCCTCGAGGTGAGACGCT-3') aus der HDL-Library 4 ein PCR-Fragment von 685 bp Länge amplifiziert. Der dritte Genomic Walking
10 Ansatz führte unter Einsatz des genspezifischen Primers DEL5-GSP1 (5'-GGTGGATGTGACGGGCGCGTACG-3') und des „nested“ genspezifischen Primers C5K-GSP1 (5'-GGTATGCCGTGGTCCAGAAGGC-3') zur Klonierung eines 924 bp PCR-Fragments aus der HDL-Library 1. Insgesamt wurden durch dieses Genomic Walking-Verfahren 2100 bp der 3' von Phagenklon P17 gelegenen genomischen hTC-Region identifiziert (s. Fig. 7).

15 Die subklonierten Fragmente sowie die Genomic Walking-Produkte wurden einzelsträngig sequenziert. Unter Verwendung der Lasergene Biocomputing Software (DNASTAR Inc. Madison, Wisconsin, USA) wurden überlappende Bereiche identifiziert und Contigs gebildet. Insgesamt wurden aus den gesammelten
20 Sequenzen der Phagenklone P12, P17, P2, P3 und P5 sowie den Sequenzdaten aus dem Genomic Walking 2 große Contigs zusammengestellt. Contig 1 besteht aus Sequenzdaten von Phagenklon P12, P17 und den Sequenzdaten aus dem Genomic Walking. Contig 2 wurde aus den Sequenzen von Phagenklon P2, P3 und P5 zusammengesetzt. Überlappende Phagenklonbereiche sind in Fig. 7 schematisch
25 dargestellt. Die Sequenzdaten der 2 Contigs sind nachfolgend dargestellt. Das ATG Startcodon in Contig 1 ist unterstrichen. Das TGA Stopcodon ist in Contig 2 unterstrichen.

Contig1:

ACTTGAGCCC AAGAGTTC AA GGCTACGGTG AGCCATGATT GCAACACCAC ACGCCAGCCT TGGTGACAGA 70
 ATGAGAGCCCT GTCTCAAAAA AAAAAAATAA AATTGAAATA ATATAAGCA TCTTCTCTGG CCACAGTGA 140
 5 ACAGAAACCG AAATCAACAA CAAGAGGAAT TTTGAAAAC ATACAAACAC ATGAAATTA AACATATAC 210
 TTCTGAATGA CCGATGAGTC AATGAAGAAA TTAAGAAAGG AATTGAAAAA TTTATTTAAG CAATGATAA 280
 CGGAAACATA ACCTCTCAAA ACCCACGGA TACAGCAAAA GCAGTGCTAA GAAGGAAGTT TATAGCTATA 350
 AGCAGCTACA TCAAAAAAGT AGAAAAGCCA GCGCAGTGG CTCATGCCCTG TAATCCCAGC ACTTTGGGAG 420
 10 GCCAAGGCGG GCAGATCGCC TGAGGTGAGG AGTTCGAGAC CAGCCTGACC AACACAGAGA AACCTTGTGC 490
 CTACTAAAAA TACAAAATTA GCTGGGCATG GTGGCACATG CCTGTAATCC CAGCTACTCG GGAGGCTGAG 560
 GCAGGATAAC CGCTTGAAAC CAGGAGGTGG AGGTTCGGGT GAGCCGGGAT TGGCCCATTT GACTCCAGCC 630
 TGGGTAAACA GAGTGAAACC CTGTCTCAAG AAAAAAATAA AAGTAGAAAA ACTTAAAAAT ACACCTTAAT 700
 GATGCACCTT AAAGAACTAG AAAAGCAAGA GCAAACTAAA CCTAAAATTG GTAAAGAAAA AGAAATAATA 770
 15 AAGATCAGAG CAGAAATAAA TGAAACTGAA AGATAACAAT ACAAAAGATC AACAAAATTA AAAGTTGGTT 840
 TTTTGAAAAG ATAAACAAAA TTGACAAACC TTTGCCAGA CTAAGAAAAA AGGAAAGAG ACCTAAATAA 910
 ATAAAGTCAG AGATGAAAAA AGAGACATTA CAATGATAC CACAGAAATT CAAAGGATCA CTAGAGGCTA 980
 CTATGAGCAA CTGTACACTA ATAAATTGAA AAACCTAGAA AAAATAGATA AATTCCTAGA TGCATACAAC 1050
 CTACCAAGAT TGAACCATGA AGAAATCCAA AGCCCAACAA GACCAATAAC AATAATGGGA TTAAGCCAT 1120
 20 AATAAAAAAGT CTCCTAGCAA AGAGAAAGCC AGGACCAAT GGCTTCCTCG CTGGATTTTA CCAATCATTT 1190
 AAAGAAGAT GAATTCAT CTACTCAAA CTATTCTGAA AAATAGAGGA AAGAATACTT CCAAACTCAT 1260
 TCTACATGCC CAGTATTACC CTGATTCCAA AACACAGCAA AACACATCA AAAACAAACA AACAAAAAAA 1330
 CAGAAAGAAA GAAACTACA GGCCAATATC CCTGATGAAT ACTGATACAA AATCCTCAA CAAAACATA 1400
 GCAGAACAAA TTAACACAA CCTTCGAAAG ATCATTCTAT GTGATCAAGT GGGATTTATT CCAGGGATGG 1470
 25 AAGGATGGTT CAACATATGC AAATCAATCA ATGTGATACA TCATCCCAAC AAAATGAAGT ACAGAACTA 1540
 TATGATTATT TCACTTTATG CAGAAAAGC ATTTGTGATA ATTCTGCACC CTTCATGATA AAAACCTCTA 1610
 AAAAACCAGG TATACAGAA ACATACAGCC CAGGCACAGT GGCTCACACC TGGCATCCCA GCACTCTGGG 1680
 AGGCCAAGGT GGGATGATT CTGGGCCCA GGAGTTTGAG ACTAGCCTGG GCAACAAAT GAGACTGTGT 1750
 CTACAAAAA CTTTTTAA AAATTAGCCA GGCATGATGG CATATGCCCTG TAGTCCAGC TAGTCTGGAG 1820
 GCTGAGGTGG GAGAACTACT TAAGCCTAGG AGGTGAGGC TGCAGTGAGC CATGAACATG TCACTGTACT 1890
 30 CCAGCCTAGA CAACAGAAACA AGACCCCACT GAATAAGAA AGGAGAGAG AGAAGGGAGA AGGAGGGAG 1960
 AAGGGAGGAG GAGGAGAGAG AGGAGGTGGA GGAGAAAGTGG AAGGGGAAGG GGAAGGGAAA GAGGAGAGAG 2030
 AAGAAACATA TTTCAACATA ATAAAGCCCT TATATGACAG ACCGAGGTAG TATTATGAGG AAAAAGCTGA 2100
 AGCCTTTCTT CTAAGATCTG GAAATGACA AGGGCCCACT TTCACCACTG TGATTCACAA TAGTACTAGA 2170
 35 AGTCCTAGCT AGAGCAATCA GATAAGAGAA AGAAATAAAA GGCATCCAAA CTGGAAAGGA AGAAGTCAAA 2240
 TTATCTGTGT TGCAGATGAT ATGATCTTAT ATCTGGAAAA GACTTAAGAC ACCACTAAAA AACTATTAGA 2310
 GCTGAAATTT GGTACAGCAG GATACAAAAT CAATGTACAA AAATCAGTAG TATTTCTATA TTCCACAGC 2380
 AAACAACTCTG AAAAAGAAAC CAAAAAGCCA GCTACAAATA AAATTAACAA GCTAGGAATT AACCAAGAA 2450
 GTGAAAGATC TCTACATGA AAATATAAAA ATGTTGATAA AAGAAATGGA AGAGGGCACA AAAAAAGAAA 2520
 40 AGATATTTCA TGTTCATGA TTGGAAGAA ATACTGTGT AAAATGTCCA TACTACCCAA AGCAATTTAC 2590
 AAATTCATG CAATCCCTAT TAAATACTA ATGACGTTCT TCACAGAAAT AGAAGAAACA ATTTAAGAT 2660
 TTGTACAGAA CCACAAAGAA CCCAGAAATG CCAAGCTAT CCTGACCAA AAGAACAAA CTGGAAGCAT 2730
 CACATTACCT GACTTCAAT TATACTACAA AGCTATAGTA ACCCAAACTA CATGGTACTG GCATAAAAAC 2800
 AGATGAGACA TGGACAGAG GAACAGAAATA GAGAATCCAG AAACAAATCC ATGCATCTAC AGTGAATCA 2870
 45 TTTTGACAA AGGTGCCAAG AACATACTT GGGGAAAAGA TAATCTCTTC AATAAATGGT CTGGAGGAA 2940
 CTGGATATCC ATATGCAAAA TAACATACT AGAACTCTGT CTCTACCAT ATACAAAGC AAATCAAAAT 3010
 GGTGAAAGG CTTAAATCTA AAACCTCAAA CTTTGCAACT ACTAAAGAA AACACCGGAG AAACCTCTCA 3080
 GGACATTGGA GTGGGCAAG ACTTCTTGAG TAATCCCTG CAGGCACAGG CAACCAAGC AAAACAGAC 3150
 AAATGGGATC ATATCAAGTT AAAAGGCTTC TGCCAGCAA AGGAAACAT CAACAAAGAG AAGAGACAAC 3220
 50 CCACAGAAAG GGAGAAATAA TTTGCAAACT ATTCATCTAA CAAGGAATTA ATACCACTA TATATAAGGA 3290
 GCTCAAACTA CTCTATAAGA AAAACACCTA ATAAGCTGAT TTTCAAAAT AAGCAAAAG TCTGGGTAGA 3360
 CATTTCTCAA AATAAGTCAT ACAAATGGCA AACAGGCATC TGAATATGTG CTCACACCA CTGATCATCA 3430
 GAGAAATGCA AATCAAACT ACTATGAGAG ATCATCTCAT CCCAGTTAAA ATGGCTTTTA TTTCAAGAC 3500
 AGGCAATAAC AAATGCCAGT GAGGATGTGG ATAAAGGAA ACCCTTGGAC ACTGTTGGTG GGAATGGAAA 3570
 55 TTGCTACCA TATGGAGAAC AGTTTGAAAG TTCTCAAAA AACTAAAAAT AAAGCTACCA TACAGCAATC 3640
 CCATTGCTAG GTATATATCT CAAAAAGGG AATCAGTGA TCAACAAGCT ATCTCCACTC CCACATTAC 3710
 TGCAGCACTG TTCATAGCAG CCAAGGTTTG GAAGCAACCT CAGTGTCCAT CAACAGACGA ATGGAAAAAG 3780
 AAAATGTGGT GCACATACAC AATGGAGTAC TACGAGCCA TAAAAAGAA TGAGATCCTG TCAGTTGCAA 3850
 CAGCATGGGG GGCCTGGTC AGTATGTTAA GTGAAATAAG CCAGGCACAG AAAGACAAAC TTTTCATGTT 3920
 60 CTCCCTTACT TGTGGGAGCA AAAATTAATA CAATTGACAT AGAAATAGAG GAGAAATGGT GTTCTAGAGG 3990
 GGTGGGGGAC AGGGTGACTA GAGTCAACAA TAATTTATTG TATGTTTAA AATAACTAAA CCGTATGTGA 4060
 TTTATACACA TTGTATGCT GTATCAAAAT ATCTCATGTA TGCTATAGAT ATAAACCTTA CTATATTAAA 4130
 AATTAATAAT TTAATGGCCA GGCACGGTGG CTCTATGCCG TAATCCAGC ACTTTGGGAG GCCGAGCGGG 4200
 65 GTGGATCACC TGAGGTCAGG AGTTTGAAC CAGTCTGGCC ACCATGATGA AACCTGTCT CTACTAAAGA 4270
 TACAAAAATT AGCCAGGCGT GGTGGCACAT ACCTGTAGTC CCAACTACTC AGGAGGCTGA GACAGGAGAA 4340
 TTGCTTGAAC CTGGGAGGCG GAGGTTGCAG TGAGCCGAGA TCATGCCACT GCACTGCAGC CTGGGTGACA 4410
 GAGCAAGACT CCATCTCAAA ACAAAACAA AAAAAAGAG ATTAATAATG TAATTTTAT GTACCGTATA 4480
 70 AATATATACT CTACTATATT AGAAGTTAAA AATTAAAAACA ATTATAAAG GTAATTAACC ACTTAATCTA 4550
 AAATAAGAAC AATGTATGTG GGGTTCTAG CTCTGAAAG AGTAAAGAT ATGGCCACGA TGGCAGAAAT 4620
 GTGAGGAGGG AACAGTGAA GTTACTGTTG TTAGACGCT ATACTCTCTG TAAGTGACTT AATTTAACC 4690
 AAAGACAGGC TGGGAGAGT TAAAGAGGCA TTCTATAAGC CCTAAACAA CTGCTAATAA TGGTGAAGG 4760
 TAATCTCTAT TAATTACCAA TAATTACAGA TATCTCTAAA ATCGAGCTGC AGAATTGGCA CGTCTGATCA 4830
 CACCGTCTCT TCATTACGAG TGCTTTTTT CTTGTGTGCT TGGAGATTTT CGATTGTGT TCGGTTTGTG 4900
 75 GTTAAACTTA ATCTGTATGA ATCTGAAAC GAAAAATGGT GGTGATTTC TCCAGAGAA TTAGAGTACC 5040
 TGGCAGGAAG .CAGGTGGCTG TGTGGACCTG AGCCACTTCA ATCTTCAAG GTCTCTGGCC AAGACCCAGG 5110

5 TGCAAGGCAG AGGCCTGATG ACCCGAGGAC AGGAAAGCTC GGATGGGAAG GGGCGATGAG AAGCCTGCCT 5180
 CGTTGGTGAG CAGCGCATGA AGTGCCCTTA TTTACGCTTT GCAAGAGATT CTCTGGATAC CATCTGGAAA 5250
 AGGCGGCCAG CGGGAATGCA AGGAGTCAGA AGCCTCCTGC TCAAACCCAG GCCAGCAGCT ATGGCGCCCA 5320
 CCCGGGCGTG TGCCAGAGGG AGAGGAGTCA AGGCACCTCG AAGTATGGCT TAAATCTTTT TTTCACCTGA 5390
 AGCAGTGACC AAGGTGTATT CTGAGGGAAG CTTGAGTTAG GTGCCCTTCT TAAACAGAA AGTCATGGAA 5460
 GCACCTTCT CAAGGGA AAA CAGAGCGCC GCTCTGCGGT CATTTACCTC TTCTCTCTCT CCTCTCTTG 5530
 CCCTCGCGGT TTCTGATCGG GACAGAGTGA CCCCCGTGGA GCTTCTCGA GCCCGTGCTG AGGACCCCTCT 5600
 TGCAAAAGGC TCCACAGACC CCCGCCCTGG AGAGAGGAGT CTGAGCCTGG CTTAATAACA AACTGGGATG 5670
 TGGCTGGGG CGGACAGCGA CGGCGGGATT CAAAGACTTA ATTCATGAG TAAATTCAC CTTTCCACAT 5740
 10 CCGAATGGAT TTGGATTTTA TCTTAATATT TTCTTAAATT TCATCAAATA ACATTACAGG CTGCAGAAAT 5810
 CCAAGGCGT AAAACAGGAA CTGAGCTATG TTTGCCAAGG TCCAAGGACT TAATAACCAT GTTCAGAGGG 5880
 ATTTTTCGCC CTAAGTACTT TTTATTGGTT TTCATAAGGT GGCTTAGGGT GCAAGGGAAA GTACACGAGG 5950
 AGAGGCTGG GCGGCAGGGC TATGAGCAGC GCAGGGCCAC CGGGGAGAGA GTCCCCGGCC TGGGAGGCTG 6020
 15 ACAGCAGGAC CACTGACCGT CCTCCCTGGG AGCTGCCACA TTGGGCAACG CGAAGGCGGC CACGCTGCGT 6090
 ATGCTCAGG ACCCATACC GGCTTCTGGG GCCCACCCAC ACTAACCCAG GAAGTCACGG AGCTCTGAAC 6160
 CCGTGGAAC GAACATGACC CTTGCCCTGC TGCTTCCCTG GGTGGGTCAA GGGTAATGAA GTGGTGTGCA 6230
 GGAATGGCC ATGTAAATTA CACGACTCTG CTGATGGGGA CCGTTCCTCT CATCATTTAT CATCTTACC 6300
 CCCAAGGACT GAATGATTCC AGCAACTTCT TCGGTTGTGA CAAGCCATGA CAAACTCAG TACAACACC 6370
 20 ACTCTTTTAC TAGGCCACA GAGCAGGSC CACACCCCTG ATATATTAAG AGTCCAGGAG AGATGAGGCT 6440
 GCTTTCAGCC ACCAGGCTGG GGTGACAACA CGCGCTGAAC AGTCTGTTC TCTAGACTAG TAGACCTTGG 6510
 CAGGCAGTCC CCCAGATTCT AGGCGCTGGT TGCTGTCTCC CGAGGGCGCC ATCTGCCCTG GAGACTCAGC 6580
 CTGGGGTGCC ACACAGGAGC CAGCCCTGTC TCCACCCCTC CGCCTCAGCT GCCTCAGCTT TCCAGCAGC 6650
 TTCTTAAAC CTGGGTGGGC CGTGTTCAG CGCTACTGTC TCACCTGTCC CACTGTGTCT TGTCTCAGCG 6720
 25 AGTAGCTCG CACGGTTCTT CCTCAGATGG GGTGTCTGTC TCCTTCCCA CACTCAGAT GCGTTGAAAG 6790
 GAGGAGATTC TCGCCTCC CAGCTGGCTC CTCTGAGCCT GAACCTGGCT CGTGGCCCC GATGAGGTT 6860
 CCTGGCGTCC GGCTGCACGC TGACCTCCAT TTCCAGGCGC TCCCGCTCTC CTGTCTATCT CGGGGCTG 6930
 CCGGTGTGTT CTTCTGTTT TGCTGCTCTT TCCACGTCCA GCTGCGTGTG TCTCTGCCCC CTAGGGTCTC 7000
 GGGGTTTTTA TAGGCATAGG ACGGGGCGGT GGTGGGCGAG GGCGCTCTTG GGAATGCAA CATTTGGGTG 7070
 30 TGAAAGTAG AGTGCCCTGTC CTCACCTAGG TCCACGGGCA CAGGCTGGG GATGGAGGCC CGCCAGGGA 7140
 CCGGCGCTTC TCTGCCAGC ACTTCTCTGC CCCCCTCCT CTGGAACACA GAGTGGCAGT TCCACAGC 7210
 ACTAAGCATC CTCTTCCCA AAGACCCAGC ATTGGCACCC CTGGACATTT GCCCACAGC CTTGGGAATT 7280
 CACGTGACTA CGCACATCAT GTACACACTC CCGTCCACGA CCGACCCCGC CTGTTTTATT TTAATAGCTA 7350
 CAAAGCAGGG AAATCCCTGC TAAATGTCC TTTAACAAC TGGTTAAACA AACGGGTCCA TCCGACGGT 7420
 35 GGACAGTTCC TCACAGTGA GAGGAACATG CCGTTTATA AGCCTGCAGG CATCTCAAGG GAATTACGCT 7490
 GAGTCAAAAC TGCCACCTCC ATGGGATACG TACGCAACAT GCTCAAAAG AAGAATTTT ACCCATGGC 7560
 AGGGGAGTGG TTAGGGGGT TAAGGACGGT GGGGGCGGCA GCTGGGGCT ACTGCACGCA CTTTACTA 7630
 AAGCCAGTTT CTTGGTCTG ATGGTATTG CTCAGTTATG GGAGACTAAC CATAGGGAG TGGGATGGG 7700
 40 GGAACCCGGA GGCTGTGCCA TCTTGGCAT CCGCGAGTGT CTTGGGCGAG ATAATGCTCT AGAGATGCC 7770
 AGGCTCTGAT TCCCCCAAC CTGTGGACAG AACCCGCCCG GCCCCAGGGC CTTGCAGGT GTGATCTCCG 7840
 TGAGGACCTT GAGGTCTGGG ATCCTTCGGG ACTACCTGCA GGCCCGAAAA GTAAATCCAGG GGTCTGGGA 7910
 AGAGGCGGG AGGAGGGTCA GAGGGGGGCA GCCTCAGGAC GATGGAGGCA GTCAGTCTGA GGTGAAAG 7980
 GAGGGAGGG CCTCGAGCCC AGGCCTGCA GCGCCTCCAG AAGCTGGA AAGCGGGGAA GGGACCCCTC 8050
 45 ACGGACCTG CAGCAGGAAG GCACGGCTG CCGTTAGCCC ACCAGGGCCC ATCTGGACC TCCGGCTCC 8120
 GTGCCATAG AGGGCACTCG CGCTGCCCTT CTAGCATGAA GTGTGTGGG ATTTGCAGAA GCAACAGGAA 8190
 ACCCATGCAC TGTGAATCTA GGATTATTTT AAAACAAAGG TTTACAGAAA CATCCAAGGA CAGGCTGAA 8260
 GTGCCCTCCG GCAAGGGCAG GGCAGGCAGC AGTGATTTTA TTTAGCTATT TTATTTTATT TACTTACTTT 8330
 CTGAGACAGA GTTATGCTCT TGTTGCCAG GCTGGAGTGC AGCGGCATGA TCTTGGCTCA CTGCAACCTC 8400
 50 CGTCTCTGG GTTCAAGCAA TTCTGTGCC TCAGCCTCCC AAGTAGCTGG GATTTACAGG GATTCACACC 8470
 ACACCCGGCT AATTTGTAT TTTTAGTAGA GATGGGCTT CACCATGTTG GTCAAGCTGA TCTCAAATC 8540
 CTGACCTCAG GTGATCCGCC CACCTCAGCC TCCCAAAGT CTGGGATTAC AGGCATGAGC CACTGCACCT 8610
 GGCTCATTTA ACCATTITAA AACTTCCCTG GGCTCAAGTC ACACCCACTG GTAAGGAGTT CATGGAGTTC 8680
 AATTTCCCTT TTAATCAGGA GTTACCCTCC TTTGATATT TCTGTAATTC TTCGTAGACT GGGGATACAC 8750
 55 CGTCTCTGA CATATTACA GTTCTGTGA CCACCTGTTA TCCCATGGGA CCCACTGCAG GGGCAGCTGG 8820
 GAGGCTGCAG GCTTCAGGTC CAGTGGGGT TGCCATCTGC CAGTAGAAAC CTGATGAGA ATCAGGGCGC 8890
 AAGTGTGGAC ACTGTCTGA ATCTCAATGT CTCAGTGTGT GCTGAAACAT GTAGAAATTA AAGTCCATCC 8960
 CTCTACTCT ACTGGATTG AGCCCTTTC CTATCCCCC CAGGGGCGAG AGGAGTTCTT CTCACTCCTG 9030
 TGGAGGAAG AATGATACTT TGTTATTTT CACTGCTGGT ACTGAATCCA CTGTTTCATT TGTTGGTTTG 9100
 60 TTTGTTTGT TTTGAGAGGC GGTTCACCT TTGTTGCTCA GGCTGGAGGG AGTGCAATGG CGCGATCTTG 9170
 GCTTACTGCA GCCTCTGCCT CCCAGGTTCA AGTGATTCTC CTGCTTCCGC CTCCCATTTG GCTGGGATTA 9240
 CAGGACCCG CCACCATGCC CAGCTAATTT TTTGATTTT TAGTAGAGAC GGGGGTGGGT GGGGTTCAAC 9310
 ATGTTGGCCA GGCTGGTCTC GAATTTCTGA CCTCAGATGA TCCACCTGCC TCTGCTCTCT AAGTGTCTG 9380
 GATTACAGGT GTGAGCCACC ATGCCAGCT CAGAATTTAC TCTGTTTGA AACATCTGGG TCTGAGGTAG 9450
 65 GAAGCTCACC CCACCTCAAG GTTGTGGTGT TTTAAGCCAA TGATAGAATT TTTTATTGT TGTTAGAACA 9520
 CTCTTGATGT TTTACATGT GATGACTAAG ACATCATCAG CTTTCAAG ACACACTAAC TGCACCCATA 9590
 ATACTGGGT GTCTTCTGGG TATCAGCAAT CTTCATTGAA TGCCGGGAGG CGTTTCTCG CCATGCACAT 9660
 GGTGTTAAT ACTCCAGAT AATCTTCTGC TTCCATTCT TCTTCTCCCT CTTTAAAAAT TGTTTCTTCT 9730
 70 ATGTTGGCT CTCTGCAGAG AACCACTGTA AGCTACAAC TAACTTTTGT TGGAAACAAAT TTCCAAACC 9800
 GCCCCTTTC CCTAGTGGA GAGACAATC ACAAACACAG CCGTTAAAAA AGGCTTAGGG ATCACTAAGG 9870
 GGATTTCTAG AAGAGCGACC TGAATCCTA AGTATTTACA AGACGAGGCT AACCTCCAG GAGCGTGACA 9940
 75 GCGGAGGAG GGTGCGAGGC CTGTTCAAA GCTAGCTCCA TAAATAAAGC AATTTCTCC GGCAGTTTCT 10010
 GAAAGTAGGA AAGGTTACAT TTAAGGTTGC GTTGTGTTAG ATTTCAAGT TTGCCGACCT CAGCTACAGC 10080
 ATCCCTGCAA GGCTCTGGGA GACCCAGAAG TTTCTGCCCC CTTAGATCC AAACCTGAGC AACCCGAGT 10150
 CTGATTCTCT GGAAGTCTCT CAGCTGTCTC GCGGTTGTGC CGGGGCCCA GGTCTGGAGG GGACAGTGG 10220
 CCGTGTGGCT TCTACTGCTG GGCTGGAAGT CGGGCTCCT AGCTCTGAG AGCTCTGAG ATGTTGGCTT CATCTGCGAG 10290
 GGCTGGACCC CGAGGCTGCC CTCCACCTG TGCGGGCGGG ATGTGACCAG ATGTTGGCTT CATCTGCGAG 10360
 ACAGAGTGCC GGGGCCCAGG GTCAAGGCCG TGTGGCTGG TGTGAGGCG CCGGTGCGG GCCAGCAGGA 10430
 GCGCCTGGCT CCATTTCCCA CCGTTTCTCG ACGGACCGC CCGGTGGGT GATTAACAGA TTTGGGGTGG 10500

	TTTGCTCATG	GTGGGGACCC	CTCGCCGCCT	GAGAACCTGC	AAAGAGAAAT	GACGGGCGTG	TGTCAAGGAG	10570
	CCCAAGTCGC	GGGGAAGTGT	TGCAGGGAGG	CACTCCGGGA	GGTCCCGCGT	GCCCGTCCAG	GGAGCAATGC	10640
	GTCTCTGGGT	TCGTCCCCAG	CCGCGTCTAC	GCGCCTCCGT	CCTCCCCCTC	ACGTCCGGCA	TTCGTGGTGC	10710
	CCGGAGCCCG	ACGCCCCCGG	TCCGGACCTG	GAGGCAGCCC	TGGGTCTCCG	GATCAGGCCA	GCGGCCAAAG	10780
5	GGTCGCCGCA	CGCACCTGTT	CCAGGGCCCT	CCACATCATG	GCCCCCTCCT	CGGGTTACCC	CACAGCCTAG	10850
	GCCGATTGCA	CCTCTCTCCG	CTGGGGCCCT	CGCTGGCGTC	CCTGCACCTT	GGGAGCGCGA	GCGGCCGCGG	10920
	GCGGGGGAAG	GCGGGGCCAG	ACCCCGGGT	CGGCCCGGAG	CAGCTGCGCT	GTCGGGGCCA	GCGCGGCTCT	10990
	CCAGTGGATT	GCGGGGCACA	GACGCCCAGG	ACCGCGCTCC	CCACGTGGCG	GAGGGAAGTG	GGACCCGGGG	11060
	ACCCGTCTCT	CCCTTTCACC	TTCAGCTCC	GCCTCCTCCG	CGCGGACCCC	GCCCCGTCCC	GACCCCTCCC	11130
10	GGGTCCCCGG	CCCAGCCCCC	TCCGGGCCCT	CCCAGCCCCC	CCCTTTCCTT	TCCGCGGCCC	GCCCCCTCTC	11200
	TGCGGGCGCG	AGTTTCAGGC	AGCGCTGCGT	CCTGCTGCGC	ACGTGGGAAG	CCCTGGCCCC	GGCCACCCCC	11270
	GCGATGCGCG	GCGCTCCCGG	CTGCCGAGCC	GTGCGCTCCC	TGCTGCGCAG	CCACTACCGC	GAGGTGCTGC	11340
	CGCTGGCCAC	GTTGCTGCGG	CGCCTGGGGC	CCAGGGCGTG	GCGGTGCTGT	CAGCGCGGGG	ACCCGGCGCG	11410
15	TTTCCGCGCG	CTGGTGGCCC	AGTGCTGGT	GTGCTGCCCC	TGGGACGCAC	GCGCGCCCCC	GCGCGGCTCT	11480
	TCCTTCCGCC	AGGTGGGCTT	CCCCGGGCTC	GGCGTCCGGC	TGGGTTGAG	GCGCGCCGGG	GGGAACCAGC	11550
	GACATCCGGA	GAGCAGCGCA	GGCGACTCAG	GGCGCTTCCC	CCGCAGGTGT	CCTGCCTGAA	GGAGCTGGTG	11620
	GCCCCAGTGC	TGCAGAGGCT	GTGCGAGCGC	GGCGCGAAGA	ACGTGCTGGC	CTTCGGCTTC	GCGCTGCTGG	11690
	ACGGGGCCCG	GCGGGGCCCC	CCCCAGGCCT	TCACACCAGC	CGTGGCGAGC	TACCTGCCCA	ACACGGTGAC	11760
	CGACGCACAT	GCGGGGAGCG	GGGCGTGGGG	GCTGCTGCTG	GCGCGCGTGG	GCGAGCAGCT	GCTGGTTTCA	11830
20	CTGCTGGCAC	GCTGCGCGCT	CTTTGTGCTG	GTGCTGCCCA	GCTGCGCCTA	CCAGGTGTGC	GGGCGCGCGC	11900
	TGTACCAAGCT	CGGCGCTGCC	ACTCAGGCCC	GGCCCCCGCC	ACACGCTAGT	GGACCCCGAA	GGCGCTGGGG	11970
	ATCGGAACGG	GCTTGGAAAC	ATAGCGTCAG	GGAGGCCGGG	GTCCCCCTGG	GCTTGGCAGC	CCCGGGTGGC	12040
	AGGAGGCGCG	GGGGCAGTGC	CAGCCGAAGT	CTGCGCTTGC	CCAAGAGGCC	CAGGCGTGGC	GCTGCCCTTG	12110
	AGCCGGAGCG	GACGCCCGTT	GGGCAGGGGT	CCTGGGCCCA	CCCGGGCAGG	ACGCGTGGAC	CGAGTGACCG	12180
25	TGCTTTCTGT	GTGGTGTAC	CTGCCAGACC	GCGCGAAGAA	GCCACCTCTT	TGGAGGGTGC	GCTCTCTGGC	12250
	ACGCGCCACT	CCCACCCATC	CGTGGGCCCG	CAGCACACAG	CAGGCCCCCC	ATCCACATCG	GCGCCACCCAC	12320
	GTCCCTGGGA	CACGCTTGT	CCCCCGGTGT	ACGCCGAGAG	CAAGCACTTC	CTCTACTCCT	CAGCGACAAA	12390
	GGAGCAGCTG	CGGCCCTCCT	TCCTACTCAG	CTCTCTGAGG	CCCAGCCTGA	CTGGCGCTCG	GAGGCTCGTG	12460
30	GAGACCATCT	TTCTGGGTTT	CAGGCCCTGG	ATGCCAGGGA	CTCCCCGAG	GTTCGCCCGC	CTGCCCCAGC	12530
	GCTACTGGCA	AATGCGGCC	CTGTTTCTGG	AGCTGCTTGG	GAACCAACGC	CAGTGGCCCT	ACGGGGTGCT	12600
	CCTCAAGACG	CAGTGGCCCG	TGCGAGCTGC	GCTCACCCCA	GCAGCCGGTG	TCTGTGCCCG	GGAGAAGCCC	12670
	CAGGGCTCTG	TGGCGGCCCC	CGAGGAGGAG	GACACAGACC	CCCGTCCGCT	GCTGCACTG	CTCCGCCAGC	12740
	ACAGCAGCCC	CTGGCAGGTG	TACGGCTTCG	TGCGGGCCCT	CCTGCGCCCG	CTGGTGCCCC	CAGGCCCTCT	12810
35	GGGCTCCAGG	CACAACGAAC	GCCGCTTCTT	CAGGAACACC	AAGAAGTTCA	TCTCCCTGGG	GAAGCATGCC	12880
	AAAGCTCTCG	TGCAGGAGCT	GACGTGGGAG	ATGAGCGTGC	GGGACTGCGC	TGGGCTGCGC	AGGAGCCAGC	12950
	GTGAGGAGGT	GCTGGCCGCT	GAGGGGCCAG	GCCCCAGAGC	TGAATGCAGT	AGGGGCTCAG	AAAGGGGGGC	13020
	AGGCAGAGCC	CTGCTCTCCT	TGCTCCATC	GTCACGTGGG	CACACGTGGC	TTTTCGCTCA	GGACGTGCGG	13090
	TGGACACGGT	GATCTCTGCC	TCTGCTCTCC	CTCCTGTCCA	GTTTGATATA	ACTTACGAGG	TTCACCTTCA	13160
40	CGTTTTGATG	GACACGCGGT	TTCCAGGCGC	CGAGGCCAGA	GCAGTGAACA	GAGGAGGCTG	GGCGCGCGAG	13230
	TGGAGCCGGG	TGCGCGCAAA	TGGGGAGAAG	TGCTTGGGAG	CACAGACGCT	CTGGCGAGGG	TGCTGCGAGG	13300
	TTACCTATAA	TCCTCTTCGC	AATTTCAAGG	GTGGGAATGA	GAGGTGGGGA	CGAGAACCCT	CTTCTCTGCG	13370
	GGGTGGGAGG	TAAGGGTTTT	GCAGGTGCAC	GTGGTCAGCC	AATATGCAGG	TTTGTGTTTA	AGATTTAATT	13440
	GTGTGTTGAC	GGCCAGGTGC	GTTGGCTCAC	GCCGGTAATC	CCAGCACTTT	GGGAAGCTGA	GGCAGGTGGA	13510
45	TCACCTGAGG	TCAGGAGTTT	GAGACCAACC	TGACCAACAT	GGTGAAACCC	TATCTGTACT	AAAAATACAA	13580
	AAATTAGCTG	GGCATTGGTG	TGTGTGCCCT	TAATCCACAG	TACTTGGGAG	GCTGAGGCGC	GAGAATCACT	13650
	TGAACCCAGG	AGGCGGAGGC	TGCAGTGAGC	TGAGATTGTG	CCATTGTACT	CCAGCCTGGG	GCAACAAGAGT	13720
	GAAACTCTGT	CTTTAAAAAA	AAAAAGTGT	CGTTGATTGT	GCCAGGACAG	GATAGAGGGA	GGGAGATAAG	13790
	ACTGTCTCTC	AGCACAGATC	CTGGTCCCAT	CTTTAGGTAT	GAAGAGGGCC	ACATGGGAGC	AGAGGACAGC	13860
50	AGATTGGTCC	ACCTGCTGAG	GAAGGGACAG	TGTTGTGGGG	TGTTAGGGGG	ATGGTGCTGC	TGGGCCCTGC	13930
	CGTGTCCTCA	CCCTGTTTTT	CTGGATTGGA	TGTTAGGAA	CCTCCGCTCC	AGCCCCCTTT	TGGTCCCGAG	14000
	TGCTCCACAG	CCCTACCGTG	GCAGCTAGAA	GAAGTCCCGA	TTTCACCCCT	TCCCCACAAA	CTCCCAAGAC	14070
	ATGTAAGACT	TCCGGCCATG	CAGACAAAGG	GGGTGACCTT	CTTGGGGCTC	TTTTTTTTCT	TTTTTCTTCT	14140
	TTATGGTGGC	AAAAGTCATA	TAACATGAGA	TTGGCACTCC	TAACACCGTT	TTCTGTGTAC	AGTGCAGAAT	14210
55	TGCTAACTCG	GCGGTGTTTA	CAGCAGGTTC	CTTGAAATGC	TGCGTCTTGC	GTGACTGGAA	GTCCCTACCC	14280
	ATCGAACCGC	AGCTGCCTCA	CACCTGCTGC	GGCTCAGGTG	GACCAACGCG	AGTCAGATAA	GCGTCATGCA	14350
	ACCCAGTTTT	GCTTTTTGTG	CTCCAGCTTC	CTTCGTTGAG	GAGAGTTTGA	GTTCTCTGAT	CAGGACTCTG	14420
	CCTGTCAATT	CTGTCTCTGT	ACTTCAGATG	AGGTCACAAT	CTGCCCCCTG	CTTATGCAGG	GAGTGAGGCG	14490
	TGGTCCCCCG	GTGTCCCTGT	CACGTGCAGG	GTGAGTGAGG	CGTTGCCCCC	AGGTGTCCCT	GTACAGTGTA	14560
60	GGGTGAGTGA	GGCGCGGCC	CCGGGTGTCC	CTGTCCCGTG	CAGCGTGATT	GAGGTGTGGC	CCCCGGGTGT	14630
	CCCTGTACAG	TGTAGGGTGA	GTGAGGCGCC	ATCCCCGGGT	GTCCCTGTCA	CGTGTAGGGT	GAGTGAGGCG	14700
	TGGTCCCCCG	GTGTCCCTGT	CCCGTGCAGG	GTGAGTGAGG	CACGTGCCCC	GGGTGTCCCT	GTACAGTGCA	14770
	GGGTGAGTGA	GGCGCGGTCC	CTCTCAGGTG	TAGGGTGAGT	GAAGGCGCGC	CCGAGGGTGT	14840	
	CCCTGTACAG	TGTAGGGTGA	GTGAGGCAAC	GTCCCTGGGT	GTCCCTCCCA	GGTATAGGGT	GAGTGAGGCA	14910
65	CTGTCCCCCG	GTGTCCCTGT	CACGTGCAGG	GTGAGTGAGG	GCGGGCCCCC	GGGTGTCCCT	CTCAGGTGCA	14980
	GGGTGAGTGA	GGCGCTGTCC	CTGGGTGTCC	CTGTCTCGTG	TAGGGTGAGT	GAAGGCTCTG	CCCCAGGTGT	15050
	CCTTGGCGTT	TGCTCACTTG	AGCTTGTCTC	TGAATGTTTG	CTCTTTCTAT	AGCCACAGCT	GCGCGGGTTG	15120
	CCCATTTGCT	GGGTAGATGG	TGCAGGCGCA	GTGCTGGTCC	CCAAGCCTAT	CTTTTCTGAT	GCTCGGCTCT	15190
	TCCTGGTCA	CTCTCCGTTT	CATTTTGCTA	CGGGGACACG	GGAGTGCAGG	CTCTCGCTCT	CCGCGTGCCA	15260
70	GGCACTGCAG	CCACAGCTTC	AGGTCCGCTT	GCCTCTGTTG	GGCCTGGCTT	GCTCACACAG	TGCGGCCAC	15330
	ATGCATGCTG	CCAATACTCC	TCTCCAGCTT	TGCTCTATGC	CGAGGCTGGA	CTCTGGGCTG	CCTGTGCTGT	15400
	CTGCCACGTG	TTGCTGGAGA	CATCCAGAA	AGGGTCTCTT	GTGCCCTGAA	GGAAAGCAAG	TCACCCAGC	15470
	CCCTCACTTT	GTCTGTTTTT	CTCCCAAGCT	GCCCCCTCTG	TTGGCCCCCT	TGGGTGGGTT	GCAACGCTTG	15540
	TCACCTTATT	CTGGGCACCT	GCCGCTCAT	GCTTAGGCTG	GGCTCTGCTT	CCAGTCCGCC	CCTCACATGG	15610
75	ATTGACGTCC	AGCCACAGGT	TGGAGTGTCT	CTGTCTGTCT	CTGTCTGTA	GACCCACGTG	GAGGCGCGG	15680
	GTTCTCGCCA	GCCTTCGTCA	GACTTCCCTC	TTGGGTCTTA	GTTTGAATTT	TCACTGATTT	ACCTCTGACG	15750
	TTTCTATCTC	TCCATTGTAT	GCTTTTTCTT	GGTTTATTCT	TTTCTCTCTT	TTCTAGCTTC	TGAGTTTATG	15820
	CATGCCCTTC	CCTCTAAGTG	CTGCCTTACC	TGCACCCCTG	GTTTGTGATG	GAAGTAATCT	CAACATCAGC	15890

	CACTTTCAAG	TGTTCTTAAA	ATACTTCAAA	GTGTTAATAC	TTCTTTTAAAG	TATTCTTATT	CTGTGATTTT	15960
	TTTCTTTTGTG	CACGCTGTGT	TTTGACGTGA	AATCATTTTG	ATATCAGTGA	CTTTTAAAGTA	TTCTTTAGCT	16030
	TATTCCTGTGA	TTTCTTTGAG	CAGTGAGTTA	TTTGAACACT	GTTTATGTTT	AAGATATGTA	GAGTATCAAG	16100
5	ATACGATAGAG	TATTTTAAAGT	TATCATTTTA	TTATTGATTT	CTAACTCAGT	TGTGTAGTGG	TCTGTATAAT	16170
	ACCAATTATT	TGAAGTTTGC	GGAGCCTTGC	TTTGTGATCT	AGTGTGTGCA	TGGTTTCCAG	AACGTGCCAT	16240
	TGTAATTG	ACATCCTGTG	AATAGTGGGC	ATGCATGTTT	ACTATATCCA	GCTTATTAAAG	GTCCAGTGCA	16310
	AAGCTTCTGT	CTCCTTCTAG	ATGCATGAAA	TTCCAAGAAAG	GAGGCCATAG	TCCCTCACCT	GGGGGATGGG	16380
	TCTGTTTCAAT	TCTTCTCGTT	TGGTAGCATT	TATGTGAGGC	ATTGTTAGGT	GCATGCACGT	GGTAGAATTT	16450
	TTATCTTCT	GATGAGTGAA	TCTTTTGGAG	ACTTCTATGT	CTCTAGTAAT	CTAGTAATTC	TTTTTTTAAA	16520
10	TTGCTCTTAG	TACTGCCACA	CTGGGCTTCT	TTTGATTAGT	ATTTTCTGTC	TGTGCTGTTT	TTCTGGCTTT	16590
	AATTTATATA	TATATATATA	TTTTTTTTTT	TTTTGAGACA	GAGTCTTGGT	CTGTGCGCCA	GGGTGAGTGC	16660
	AGTGGTGTGA	TCACAGGTCA	GTGTAACTTT	TACCTTCTGG	CCTGAGCCGT	CCTCTCACCT	CAGCCTCCTG	16730
	AGTAGCTGGA	ACTGCAGACA	CGCACCGCTA	CACCTGGCTA	ATTTTAAAT	TTTTTCTGGA	GACAGGGTCT	16800
	TGCTGTGTTG	CCCAGGCTGG	TCTCAAACCT	TGGGACTCAA	GGGATCCATC	TACCTCGGCT	TCCCAAAGTG	16870
15	CTGAATTACA	GGCATGAGCC	ACCATGTCTG	GCCTAATTTT	CAACACTTTT	ATATCTTAT	AGTGTGGGTA	16940
	TGTCCTGTTA	ACAGCATGTA	GGTGAATTTT	CAATCCAGTC	TGACAGTCGT	TGTTTAACTG	GATAACTCTGA	17010
	TTTATTTTCA	TTTTTTTGTG	ACTAGAGACC	CGCCTGGTGC	ACTCTGATTC	TCCACTTGCC	TGTTGCGATGT	17080
	CCTCGTTCCC	TTGTTTCTCA	CCACCTCTTG	GGTTGCCATG	TGCGTTTCTT	GCCGAGTGTG	TGTTGATCCT	17150
	CTCGTTGCTT	CCTGGTCACT	GGGCATTTTG	TTTTATTCTT	CTTTGCTTAG	TGTTACCCCT	TGATCTTTTT	17220
20	ATTGTCGTTG	TTTGCTTTTG	TTTATTGAGA	CAGTCTCACT	CTGTCAACCA	GGCTGGAGTG	TAATGCGACA	17290
	ATCTCGGCTC	ACTGCAACCT	CTGGCTCCTC	GGTTCAGACA	GTTCTCATTC	CTCAACCTCA	TGAGTAGCTG	17360
	GGATTACAGG	CGCCCCACCA	CACGCTGGCC	TAATTTTGT	ATTTTATAGTA	GAGATAGGCT	TTCAACATGT	17430
	TGGCCAGGCT	GGTCTCAAA	TCCTGACCTC	AAGTGATCTG	CCCGCCTTGG	CCTCCACAG	TGCTGGGATT	17500
	ACAGGTGCAA	GCCACCGTGC	CGGCGATACC	TTGATCTTTT	AAAATGAAGT	CTGAAACATT	GTACCCCTTG	17570
25	TCCTGAGCAA	TAAGACCCCT	AGTGATTTT	AGCTCTGGCC	ACCCCCACAG	CTGTGTGCTG	TTTTCCCTGC	17640
	TGACTTAGTT	CTATCTCAGG	CATCTTGACA	CCCCCACAA	CTAAGCATT	TTAATATTGT	TTTCCGTGTT	17710
	GAGTGTTTCT	GTAGCTTTTG	CCCCGCCCTG	CTTTTCTCTC	TTTGTCTCCC	GCTGTCTTC	TGCTCAGGC	17780
	CGCGCGTCTG	GGGTCCCTTT	CTTGTCTCTT	TGCGTGGTTC	TTCTGTCTTG	TTATTGCTGG	TAAACCCAG	17850
30	CTTTACCTGT	GCTGGCCTCT	ATGGCATCTA	GCGACGTCGG	GGGACCTCTG	CTTATGATGC	ACAGATGAAG	17920
	ATGTGGAGAC	TCACGAGGAG	GGCGGTATC	TTGGCCCGTG	AGTGTCTGGA	GCACACGCTG	GCCAGCGTTC	17990
	CTTAGCCAGT	GAGTGACAGC	AACGTCCGCT	CGCGCTGGGT	TCAGCTGGGA	AAACCCACAG	CATGTGCGGG	18060
	TCTGGTGGCT	CCGCGGTGTC	GAGTTTGAAA	TCGCGCAAA	CTGCGGTGTG	CGCCACGCTC	TGACGGTGCT	18130
	GCCTGGCGGG	GGAGTGTCTG	CTTCTCCCTT	CTGTCTTGGG	AACAGGACA	AAGGATGAGG	CTCCGAGCCG	18200
	TTGTGCGCCA	ACAGGAGCAT	GACGTGAGCC	ATGTGGATAA	TTTTAAATTT	TCTAGGCTGG	CGCGCGTGGC	18270
35	TCAGCCCTGT	AATCCAGACA	CTTTGGGAGG	CCAAGCCGGG	TGGATCACGA	GGTCAGGAGG	TGAGAGCCAT	18340
	CCTGGCCAAC	ATGATGAAAC	CCCATCTGTA	CTAATAAACAC	AAAAATTAGC	TGGCGTGGT	GGCGGTGGCC	18410
	TGTAATCCCA	GCTACTCGGG	AGGCTGAGGC	AGGAGAATTG	CTTGAACCTG	GGAGTTGGAA	GTTCGAGTGA	18480
	GCCGACATTG	CACCACTGCA	CTCCAGCCTG	GCAACACAGC	GAGACTCTGT	CTCAAAAAAA	AAAAAAAAAA	18550
40	AAAAAATAAA	AATTCTAGTA	GCCACATTAA	AAAAAGTAAA	AAGAAAGGTT	GAAATTAATG	TAATAATAGA	18620
	TTTTACTGAA	GCCACGATG	TCCACACCTC	ATCATTTTAA	GGTGTATTG	GTGGGAGCAT	CACCTACAGG	18690
	ACATTTGACA	TTTTTTGAGC	TTTGTCTCGG	GGATCCCTGT	TGTAGGTCCC	GTGCGTGGCC	ATCTCGGGCT	18760
	GGACCTGCTG	GGCTTCCCAT	GGCATGCTG	GTTGTACACG	ATGGTGACAG	TCCGGGATGA	GGTCCGACGG	18830
	CCCTCAGTGA	GCTGGATGTG	CAGTGTCGGG	ATGGTGACAG	TCTGGGATGA	GGTCCGACAG	CCCTGCTGTG	18900
45	AGCTGGATGT	GTGGTGCTG	GATGGTGAGC	GTGAGGGTGG	AGGTCTCCAG	GCCCTCGGTC	AGCTGGAGGT	18970
	ATGGAGTCCG	GATGATGCA	GTCCGGGGTG	AGGTGCGCCAG	GCCCTGCTGT	GAGCTGGATG	TGTGGTGTCT	19040
	GGATGGTGCA	GGTCAGGGGT	GAGGTCTCCA	GGCCCTCGGT	AAGCTGGAGG	TATGGAGTCC	GGATGATGCA	19110
	GGTCCGGGGT	GAGGTGCGCA	GGCCCTGCTG	TGAGCTGGAT	GTGTGTGTCT	TGGATGGTGC	AGGTCTGGGG	19180
	TGAGGTGACC	AGGCCCTGCG	GTGAGCTGGG	GTGCGGGTGT	CTGGATGGTG	CAGGTCTGGA	GTGAGGTGCG	19250
50	CAGACCGTGC	CAGACCATGC	GGTGAGCTGG	ATATGCGGTG	TCCGGATGTT	GCAGGTCTGG	GGTGAGGTGG	19320
	CCAGGCCCTG	CTGTGAGTTG	GATGTGGGGT	GTCCGGATGC	TGCAGGTCCG	GTGTGAGGTC	ACCAGGCCCT	19390
	GCTGTGAGCT	GGATGTGTGG	TGTCTGGATG	GTGCAAGTCT	GGGTGAAAGG	TCCGCCAGGC	CTCGCTTGTG	19460
	AGCTGGATGT	GTGGTGTCTG	GATGGTGCA	GTCTGGAGTG	AGGTGCGCCAG	GCCCTCGGTC	AGCTGGATGT	19530
	GCAGTGTCCA	GATGGTGCA	GTCCGGGGTG	AGGTGCGCCAG	ACCCTGCGGT	GAGCTGGATG	TCCGGTGTCT	19600
55	GGATGGTGCA	GGTCTGAGAT	GAGGTGCGCA	GGCCCTCGGT	GAGCTGGATG	TATGGAGTCC	GGATGGTGCC	19670
	GGTCCGGGGT	GAGGTGCGCA	GACCTTCTGT	TGAGCTGGAT	GTGCGGTGTC	TGGATGGTAC	AGGTCTGGAG	19740
	TGAGGTGCGC	AGACCTGCT	GTGAGCTGGA	TATGCGGTGT	CCGGATGGTG	CAGGTGAGGG	GTGAGGTCTC	19810
	CAGGCCCTCG	GTGAGCTGGA	GGTATGGAGT	CCGGATGATG	CAGGTCCGGG	GTGAGGTGCG	CAGGCCCTCG	19880
	TGTGAACCTG	ATGTGCGGGC	TCTGGATGGT	GCAGGTCTGG	GGTGTGGTCT	CCAGGCCCTC	GGTGAGCTGG	19950
60	AGGTATGGAG	TCCGGATGAT	GAGGTCCGGG	GGTGAGGTCT	CCAGGCCCTG	CTGTGAGCTG	GATGTGCGGC	20020
	GTCTGGATGG	TGCAGGTCTG	GGGTGTGGTG	GCCAGGCCCT	CGGTGAGCTG	GAGGTATGGA	GTCCGGATGA	20090
	GGGTGAGGTC	GCCAGGCCCT	GCTGTGAGCT	GGATGTGCTG	TATCCGGATG	GTGAGTCCG	GTGAGTCCG	20160
	GGGTGAGGTC	GCCAGGCCCT	GCTGTGAGCT	GGATGTGCTG	TATCCGGATG	GTGAGTCCG	GTGAGTCCG	20230
	CACCAAGGCC	TGCGGTGAGC	TGTTTGTGCG	GTGTCCGGTT	GCTGCAGGTC	CGGGGTGAGT	TCCCAAGGCC	20300
65	CTCGGTGAGC	TGGATGTGCG	GTGTCCCGGT	GTCCGGATGG	TGCAGGTCCA	GGGTGAGGTC	GCTAGGCCCT	20370
	TGGTGGGCTG	GATGTGCGGT	GTCCGGATGG	TGCAGGTCTG	GGGTGAGGTC	GCCAGGCCCT	TGGTGGGCTG	20440
	GATGTGCGGT	GTGTGATGCG	TGCAGGTCTG	GGGTGAGGTC	GCCAGGCCCT	TGGTGGGCTG	GATGTGCGGT	20510
	GTCCGGATGG	TGCAGGTCTG	GGGTGAGGTC	GCCAGGCCCT	GCTGTGAGCT	GGATGTGCGG	TGCTGGATG	20580
70	GTCCGGATGG	TGCAGGTCTG	GGGTGAGGTC	GCCAGGCCCT	TTCGGTGAGC	TGGATGTGCG	GTGTCCGGAT	20650
	GGGGTGGAGG	TGCCAGGCC	CTCGGTTAG	CTGGATATGC	GGTGTCCGGA	TGGTGCAGGT	CGTCCGAGG	20720
	GTCAACAGGC	CCTGCGGTTA	GCTGGATGTC	CGGTGTCTGG	ATGGTGCAGG	TCCGGGGTGA	GGTCCGAGG	20790
	CCCTGCTGTG	AGCTGGATGT	GCTGTATCCG	GATGGTGCA	GTCCGGGGTG	AGGTGCGCAG	GCCCTGCAGT	20860
	GAGCTGGATG	TGCTGTATCC	GGATGGTGCA	GGTCTGGCGT	GAGGTGCGCA	GGCCCTGCGG	TTAGCTGGAT	20930
	ATGCGGTGTC	GGATGGTGCA	GGTCCGGGGT	GAGGTGACCA	GGCCCTGCGG	TTAGCTGGAT	GTGCGGTGTC	21000
75	CGGATGGTGC	AGGTCTGGGG	TGAGGTGCGC	AGGCCCTGCT	GTGAGCTGGA	TGTGCTGTAT	CCGGATGGTG	21070
	CAGGTCCGGG	GTGAGGTGCG	CAGGCCCTG	GGTGGATGTC	ATGTGCTGTA	TCCGGATGGT	GCAGGTCTCG	21140
	CGTGAGGTGC	CCAGGCCCTG	CGGTGAGCTG	GATGTGCACT	GTACGGATGG	TGCAGGTCCG	GGGTGAGGTC	21210
	GCCAGGCCCT	GCGGTGGGCT	GATGTGTGT	TGTCTGGATG	GTGCAAGGTC	GGGTGAGGTC	GCCAGGCCCT	21280

	TGCGGTGAGC	TGGATGTGTG	GTGTCTGGAT	GCTGCAGGTC	CGGGGTGAGT	TCGCCAGGCC	CTCGGTGAGC	21350
	TGGATATGCG	GTGTCCCGT	GTCCGAATGG	TGCAGGTCCA	GGGTGAGGTC	GCCAGGCCCT	TGGTGGGCTG	21420
	GATGTGCCGT	GTCCGGATGG	TGCAGGTCTG	GGGTGAGGTC	GCCAGGCCCT	TGGTGAAGCTG	GATGTGCCGT	21490
5	GTCCGGATGG	TGCAGGTCCG	GGGTGAGGTC	ACCAGGCCCT	CGGTGATCTG	GATGTGGCAT	GTCTTCTCG	21560
	TTTAAGGGGT	TGGCTGTGTT	CCGGCCGCGAG	AGCACCGTCT	CGGTGAGGAG	ATCTTGCCCA	AGTTCTTGCA	21630
	CTGGCTGATG	AGTGTGTACG	TCGTGAGACT	GCTCAGGTCT	TTCTTTTATG	TCACGGAGAC	CACGTTTCAA	21700
	AAGAAACAGG	TCTTTTCTA	CCGGAAGAGT	GTCTGGAGCA	AGTTGCAAG	CATTGGAATC	AGGTACTGTA	21770
	TCGCCAGGCC	AGGCCTCTGC	TTCTCGAAGT	CCTGGAACAC	CAGCCCGGCC	TCAGCATCGC	CCTGTCTCCA	21840
	CTTGCCCTGTG	CTTCCCTGGC	TGTGCAGCTC	TGGGCTGGGA	GCCAGGGGCC	CCGTACACAG	CCTGGTCCAA	21910
10	GTGGATTCTG	TGCAAGGCTC	TGACTGCCCTG	GAGCTCACGT	TCTCTTACTT	GTAATAATCAG	GAGTTTGTGC	21980
	CAAGTGGTCT	CTAGGGTTTG	TAAAGCAGAA	GGGATTTAAA	TTAGATGGAA	AACTTACCAC	TAGCTCCTT	22050
	GCCTTTCCCT	GGGATGTGGG	TCTGATTCTC	TCTCTCTTTT	TTTTTCTTTT	TTTGAGATGG	AGTCTCACTC	22120
	TGTTGCCAGG	GCTGGAGTGC	AGTGGCATAA	TCTTGGCTCA	CTGCAACCTC	CACCTCCTGG	GTTTAAGCGA	22190
	TTCAACAGCC	TCAGCCTCCT	AAGTAGCTGG	GATTACAGGC	ACCTGCCACC	ACGCTGGCT	AATTTTGTGA	22260
15	CTTTTAGGAG	AGACGGGGTT	TCACCATGTT	GGCCAGGCTG	GTCTCGAACT	CATGACCTCA	GGTGATCCAC	22330
	CCACCTTGCC	CTCCCAAAGT	GCTGGGTTTA	CAGGCTAAGC	CACCGTGCCC	AGCCCCGAT	TCTCTTTTAA	22400
	TTCATGCTGT	TCTGTATGAA	TCTTCAATCT	ATTGGATTTA	GGTCATGAGA	GGATAAAATC	CCACCCACTT	22470
	GGCGACTCAC	TGCAGGGAGC	ACCTGTGCGAG	GGAGCACCTG	GGGATAGGAG	AGTTCCACCA	TGAGCTAACT	22540
20	TCTAGGTGGC	TGCATTGTAA	TGGCTGTGAG	ATTTTGTCTG	CAATGTTCCG	CTGATGAGAG	TGTGAGATTG	22610
	TGACAGATTC	AAGCTGGATT	TGCATCAGTG	AGGGACGGGA	GGCTGGTCT	GGGAGATGCC	AGCCTGGCTG	22680
	AGCCACAGCC	ATGGTATTAG	CTTCTCCGTT	TCCCGCCAG	GCTGACTGTG	GAGGGCTTTA	GTGAGAAAGT	22750
	CAGGGCTTCC	CCAGCTCCCC	TGCACACTCG	AGTCCCTGGG	GGGCTTGTG	ACACCCCATG	CCCCAAATCA	22820
	GGATGTCTGC	AGAGGGAGCT	GGCAGCAGAC	CTGTCAGAG	GTAACACAGC	CTCTGGGCTG	GGGACCCCGA	22890
25	CGTGGTCTGC	GGGCCATTTC	CTTGATCTG	GGGGAGGGTC	AGGGCTTTCC	CTGTGGGAAC	AAGTTAATAC	22960
	ACAATGCACC	TTACTTAGAG	TTTACACGTA	TTTAATGGTG	TGCGACCCAA	CATGGTCATT	TGACCAAGAT	23030
	TTTGGAAAGA	ATTTAATTGG	GGTGACCGGA	AGGAGCAGAC	AGACGTGGTG	GTCCCAAGA	TGCTCTTTGT	23100
	CACTACTGGG	ACTGTGTTTC	TGCCTGGGGG	GCCTTGGAGG	CCCTCTCTCC	CTGGACAGGG	TACCGTGCTG	23170
	TTTCTACTCT	GCTGGGCCCTG	CGGCCCTGGG	TCAGGGCACC	AGCTCCGGAG	CACCCGCGGC	CCCACTGTCC	23240
30	ACGGAGTGCC	AGGCTGTGAG	CCACAGATGC	CCAGTCCAG	GTGTGGCCGC	TCCAGCCCCC	GTGCCCCCTT	23310
	GGGTGGTTTT	GGGGGAAAG	GCCAAGGGCA	GAGGTGTCAG	GAGACTGGTG	GGCTCATGAG	AGCTGATTCT	23380
	GCTCCTTGGC	TGAGCTGTCC	TGAGCAGCCT	CTCCCCCCT	CTCATCTGA	AGGGATGTGG	CTCTTTCTAC	23450
	CTGGGGGTCC	TGCTTGGGGC	CAGCCTTGGG	CTACCCCACT	GGCTGTACCA	GAGGGACAGG	CATCTGTGTA	23520
	GGAGGGGCAT	GGGTTACAGT	GGCCCCAGAT	CGACCTGGG	ACCAGGCTCC	CTGGTGTGTA	TGGTGGGACA	23590
35	GTCACCTTGG	GGTTTACAGC	CCGGACTGGG	CGTCCCCAGG	GTGACTATA	GGACCAAGTG	TCCAGGTGCC	23660
	CTGCAAGTAG	AGGGGCTCTC	AGAGGGCTCT	GGCTGGCATG	GGTGGACGTG	GCCCCGGGCA	TGGCCTTCAG	23730
	CGTGTGCTGC	CGTGGGTGCC	CTGAGCCCTC	ACTGAGTCCG	TGGGGGCTTG	TGGCTTCCCG	TGAGCTTCCC	23800
	CCTAGTCTGT	TGTCTGGCTG	AGCAAGCCTC	CTGAGGGGCT	CTCTATTGCA	GACAGCACTT	GAAGAGGGTG	23870
40	CAGCTGCCGG	AGCTGTCCGA	AGGCAGGTC	AGGCAGCATC	GGGAAGCCAG	GCCCCGCGTG	CTGACGTCCA	23940
	GACTCCGCTT	CATCCCCAAG	CCTGACGGGC	TGCGGCCGAT	TGTGAACATG	GACTACGTCT	TGGGAGCCAG	24010
	AACGTTTCCG	AGAGAAAAGA	GGGTGGCTGT	GCTTTGGTTT	AACCTCCTTT	TTAAACAGAA	GTGCTTTTGA	24080
	GGCCACATT	TGGTATCAGC	TTAGATGAAG	GGCCCGGAGG	AGGGGCCACG	GGACACAGCC	AGGGCCATCG	24150
	CACGGCGCCA	ACCCATTGTG	CGGCACAGTG	AGGTGGCCGA	GGTGCCGGTG	CCTCCAGAAA	AGCAGCGTGG	24220
	GGGTGTAGGG	GGAGCTCCTG	GGGCAGGGAC	AGGCTCTGAG	GACCACAAGA	AGCAGCCGGG	CCAGGGCGTG	24290
45	GATGCAGCAC	GGCCCGAGGT	CCTGGATCCG	TGTCCTGCTG	TGGTGGCGAG	CCTCCGTGGC	CTTCCGCTTA	24360
	CGGGGGCCCG	GGACACGGCC	ACGACTGCCA	GGAGCCCAAC	GGGCTCTGAG	GATCCTGGAC	CTTCCGCCAC	24430
	GGCTCCTGCA	CCCCACCCCT	TGGGCTGCGG	TGGCTGCGGT	GACCCCGTCA	TCTGAGGAGA	GTGTGGGGTG	24500
	AGGTGGACAG	AGGTGTGGCA	TGAGGATCCC	GTGTGCAACA	CACATGCGGC	CAGGAACCCG	TTTCAACACG	24570
	AGTCTGAGGA	AGCTGGGAGG	GGTTCTAGGT	CCCGGGTCTG	GCTGGCTGGG	GACACTGGGG	AGGGGCTGCT	24640
50	TCTCCCTGTG	GTCCCTATGG	TGGGGTGGGC	ACTTGGCCGG	ATCCACTTTC	CTGACTGTCT	CCCATGCTGT	24710
	CCCGGCCAGG	CCGAGCGTCT	CACCTCGAGG	GTGAAGGCAC	TGTTACAGCT	GCTCAACTAC	GAGCGGGCGC	24780
	GGCGCCCCGG	CCTCCTGGGC	GCCTCTGTGC	TGGGCTTGGG	CGATATCCAC	AGGGCCTGGC	GCACCTTCGT	24850
	GCTGCGTGTG	CGGGCCCGAG	ACCCGCCCGC	TGAGCTGTAC	TTTGTCAGG	TGGGTGCGCG	GGACCCCGGT	24920
	GAGCAGCCCT	GCTGGACCTT	GGGAGTGGCT	GCCTGATTGG	CACCTCATGT	TGGGTGGAGG	AGGTACTCCT	24990
55	GGGTGGGCGG	CAGGAGTGC	AGGTGACCCT	GTCACTGTG	AGGACACACC	TGGCACCTAG	GGTGGAGGCC	25060
	TTCAAGCCTT	CCTGAGCAC	ATGGGGCCGA	CTGTGCACCC	TGACTGCCCG	GGCTCCTATT	CCCAAGGAGG	25130
	GTCCCACTGG	ATTCCAGTTT	CCGTACAGAA	AGGAACCCGA	ACGGCTCAGC	CACCAAGGCC	CGGTGCTTTG	25200
	CACCCAGCTC	CTGAGCCAGG	GGTCTCCTGT	CCTGAGGCTC	AGAGAGGGGA	CACAGCCCGC	CCTGCCCTTG	25270
	GGGTCTGGAG	TGGTGGGGT	CAGAGAGAGA	GTGGGGGACA	CCGCCAGGCC	AGGCCCTGAG	GGCAGAGGTG	25340
60	ATGTCTGAGT	TTCTGCGTGG	CCACTGTGAG	TCTCTCGCC	TCCACTCACA	CAGGTGGATG	TGACGGGCGC	25410
	GTACGACACC	ATCCCCCAGG	ACAGGCTCAC	GGAGGTATC	GCCAGCATCA	TCAAACCCCA	GAACACGTAC	25480
	TGCGTGGCTC	GGTATGCCGT	GGTCCAGAAG	CGCCGCCATG	GGCACGTCCG	CAAGGCCCTT	AAGAGCCACG	25550
	TAAGGTTTCA	GTGTGATAGT	CGTGTCCAGG	ATGTGTGTCT	CTGGGATATG	AATGTGTCTA	GAATGCAGTC	25620
	GTGTCTGTGA	TCCGTTTCTG	TGGTGGAGGT	ACTTCCATGA	TTTACACATC	TGTGATATGC	GTGTGTGGCA	25690
65	CGTGTGTGTC	GTGTGTGATG	TATCTGTGGC	GTGCATATTT	GTGGTGTGTG	TGTGTGTGGC	ACGTGTGTGT	25760
	CCATGGTGTG	TGTGCCCTGTG	GTGTGATGTG	GTGTGTGTCT	GTGACACGTG	CATGTTTATG	CTGTGTGCTG	25830
	CATGTCTGTG	ATGTGCCCTAT	TTGTGGTGTG	TGTGTGATG	TGTCCGTGAC	ATATGCGTGT	CTATGGCATG	25900
	GGTGTGTGTG	GCCCTTGGCC	CTTACTCCTT	CCTCTCCAG	GCATGGTCCG	CACCATTTGT	CTACGCTCT	25970
	CGGGTGTGTG	TTTGGGGAGC	TCCACATTCA	GGGTCTCTAC	TTCTAGCATG	GGTGCCCTTG	TCCTGTGACA	26040
70	GGGCTGGGCC	TTGAGACTG	TAAAGCAGGT	TTGAGAGGAG	AGTAGGGATG	CTGGTGGTAC	CTTCTGGGAC	26110
	CCCTGGCCACC	CCAGGACACC	CAGTCTGGCC	TATGCCGGCT	CCATGAGATA	TAGGAAGGCT	GATTTCAGGCC	26180
	TCGCTCCCGC	GGACACACTC	CTCCAGAGCC	GGCCGGGGCC	CTTGGGGCTC	GGCAGGGGTC	AAAGGGGCCC	26250
	TGGGCTTGGG	TTCCCAACCA	GTGGTCATGA	GCACGCTGGA	GGGGTAAGCC	CTCAAAGTCG	TGCCAGGCCG	26320
	GGGTGCAGAG	GTGAAGAAGT	ATCCCTGGAG	CTTCGGTCTG	GGGAGAGGCA	CATGTGGAAA	CCCCAAGGGA	26390
75	CCTCTTCTC	TGACTTCTTG	AGCT					26414

Contig 2:

	TGTGGGATTG	TTTTTCATGT	GTGGGATAGG	TGGGGATCTG	TGGGATTGGT	TTTTATGAGT	GGGGTAACAC	70
	AGAGTTC AAG	GCGAGCTTTC	TTCCTGTAGT	GGGTCTGCAG	GTGCTCCAAC	AGCTTTATTG	AGGAGACCAT	140
5	ATCTTCTCTT	GAACATAGGT	CGGGTTTATA	GTAAGTCAGG	GGTGTGGAGG	CCTCCCTGGG	GCTCCCTGTT	210
	CTGTTTCTTC	CACTCTGGGG	TCGTGTGGTG	CCTGCTGTGG	TGTGTGGCCG	GTGGGCGAGG	CTTCCAGGCC	280
	TCCTTGTGTT	CATTGGCCTG	GATGTGGCCC	TGGCTACGCT	CCGTCTTTGG	AATTCCCTCG	CGAGTTGGAG	350
	GCTTCTTTTC	TTTCTTTTTT	TCITTTCTTT	TTTTTTTTTT	TGATAACAGA	GTCTCGCTCT	TTTTTGCCCA	420
	GGCTGGAGTG	GTTTGGCGTG	ATCTTGGCTC	ACTGCAACCT	GTGCTTCCTG	AGTTCAAGCA	ATTCTCTTGC	490
10	CTCAGCCTCC	CAAGTAGCTG	GAATTATAGG	CGCCCACCAC	CATGCTGACT	AATTTTTGTA	ATTTTAGTAG	560
	AGACGAGGTT	TCTCCATGTT	GGCCAGGCTG	GTCTCGAACT	CCTGACCTCA	GGTGATCTCT	CCACCTCGCG	630
	CTCCCAAAGT	GCTGGGATGA	CAGGTGTGAA	CCGCCGCGCC	CGGCCGAGAC	TCGCTTCTCG	CAGCTTCCGT	700
	GAGATCTGCA	GCGATAGCTG	CCTGCAGCCT	TGGTGTGTAC	AACCTCCGTT	TTCTTCTCC	AGGTCTCGCT	770
	AGGGGTCTTT	CCATTTCAATG	ACTCTCTTCA	CAGAAAGATT	TCACGTGTGC	TGATTTCCCG	GCTGTTCCCT	840
15	CGCTAATTGG	TGCTCTGCTG	TTATCGATGG	CCTCCTTCCA	TTTCTTTTAG	GCTTTGTTTA	TTGTTGTTTT	910
	TCGGCTCCTT	TGAAGGAAAA	GTTTTCGATTA	TGGATGTTTG	AACCTTCTTT	TCTAAACAAG	CATCTGAAGT	980
	TGCGGTTTTT	CCTCTAAAGC	AGGGATCCCG	AGGCCCTTGG	CTGTGGAGTG	GCACCGGTCT	GGGGCTGTTT	1050
	AGGAACCCGG	CGCACAGCGG	GAGGCTAGGT	GGGGTGTGGG	GAGCCAGCGT	TCCCGCCTGA	GGCCCGCCCC	1120
	TCTCAGATCA	CGAGTGGCAT	GCGGTGCTCA	GAGGCGCAC	CACCTTACTG	AGAACTGTGC	GTGAGAGGGG	1190
20	TCTAGATTCT	GTGCTCTTCA	TGGGAATCTA	ATGCCCTGATG	ATCTGAGGTG	GAACCGTTTG	CTCCCAAAAC	1260
	CATCCCTCTT	CCCACGTGCT	TCCTGTGGAA	AAATCGTCTT	CCACGAAACC	AGTCCCTGGT	ACCACAATGG	1330
	TTGGGGACCC	TGTGCTAAAG	ACCTGCTTCA	GCAGCCTCTC	GTCACTGTTG	ATATATTGGC	TTTTCTGTGT	1400
	TGAGTCCAGA	ATAATTACGG	ATTCTCTGTA	TGCTTTCGCG	CGACCTCAGA	CCCATGGGCT	ATTGTGGGGC	1470
	GTGTTGCCCT	CTCCTGGGTT	GGGAAGGGTG	CAGGCCCTAT	GTACCTTCTT	GTTACTGCTC	TCCAGGTTGG	1540
25	TTCTCAGGGT	TGAATCGTAC	TCGATGTGGT	TTTAGCCCAAC	GGCCTTCCCG	CCAGCTCCTG	GGGGCTGGGG	1610
	AACATGCTGA	AGCACAGAGT	CACCGTCCGC	GTCTTTTGAT	GCCTCACAA	CTCGAGGCGT	CCTGTGTCCT	1680
	TGTTAGTGTG	TGTCACGTGC	CTGCTCACAT	CTGTCTTGGG	GGACGCAGGG	GCTTAGCAGG	TCCCGTAGTA	1750
	AATGACAAGC	GTCTTGGGGG	AGTCTGCAGA	ATAGGAGGTT	GGGGTGCCGG	TCTCTCTCCC	GGCTCTTCAG	1820
	ACTCTTCTCC	TGCCCTGTGCT	GTGGCTGCAC	CTGCATCCCT	GCAATCCCTC	CAGCACTGGG	CTGGAGAGGC	1890
30	CCGGGAGCTC	GAGTGCCACT	TGTGCCACGT	GACTGTGGAT	GGCAGTCCGT	CACGGGGGTC	TGATGTGTGG	1960
	TGACTGTGGA	TGGCGGTTGG	TACACAGGGT	CTGATGTGTG	GTGACTGTGG	ATGGCGGTGC	TGGGGTCTGA	2030
	TGTGGTGACT	GTGGATGGCG	GTCTGTTGGT	CTGATGTGTG	GTGACTGTGG	ATGGCGGTGC	TGGGGTCTGA	2100
	TGTGGTGACT	GTGGATGGCG	GTCTGTTGGT	CTGATGTGTG	GTGACTGTGG	ATGGCGGTGC	TGGGGTCTGA	2170
	TGGTGACTGT	GGATGGCAGT	CGTGGGGTCT	GATGTGTGGT	GACTGTGGAT	GGCGGTCTGT	GGGTCTGATG	2240
35	TGGTGACTGT	GGATGGCAGT	CGTGGGGTCT	GATGTGTGGT	GACTGTGGAT	GGCGGTCTGT	GGGTCTGATG	2310
	TGTGGTGACT	GTGGATGGCG	GTCTGTTGGT	CTGATGTGTG	GTGACTGTGG	ATGGCGGTGC	TGGGGTCTGA	2380
	TGTGTGGTGA	CTGTGGATGG	CGGTGTTGGG	GTCTGATGTG	GTGACTGTGG	ATGGCGGTGC	TGGGGTCTGA	2450
	TGTGTGGTGA	CTGTGGATGG	TGATCGGTCA	CAGGGGTCGT	ATGTGTGGTG	ACTGTGGATG	CGCGTCTGTT	2520
40	GGTCTGATGT	GTGGTGACTG	TGGATGGTGA	TCGGTCACAG	GGGTCTGATG	TGTGGTGACT	GTGGATGGCG	2590
	GTCTGTTGGT	CTGATGTGTG	GTGACTGTGG	ATGGCGGTTG	GTCCCGGGGG	TCTGATGTGT	GGTGACTGTG	2660
	GATGGGATC	GGTCACAGGG	GTCTGATGTG	TGGTGACTGT	GGATGGCGGT	CGTGGGGTCT	GATGTGTGGT	2730
	GACTGTGGAT	GGCGGTCTGT	GGGTCTGATG	TGTGGTGACT	GTGGATGGCG	GTCTGTTGGT	CTGATGTGGT	2800
	GACTGTGGAT	GGCGGTCTGT	GGGTCTGATG	TGTGGTGACT	GGATGGCGGT	CGTGGGGTCT	GATGTGTGGT	2870
45	GACTGTGGAT	GGCGGTCTGT	GGGTCTGATG	TGTGGTGACT	GGATGGCGGT	CGTGGGGTCT	GATGTGTGGT	2940
	GTGGTGACTG	TGGATGGCAG	TCGTGGGGTC	TGATGTGTGG	TGACTGTGGA	TGGCGGTCTG	GGGTCTGATG	3010
	CTGTGGTGAC	TGTGGATGGC	GGTCTGTTGG	TCTGATGTGT	GGTGACTGTG	GATGGCGGTC	GTGGGGTCTG	3080
	ATGTGTGGTG	ACTGTGGATG	GGGTCTGTTG	GGTCTGATGT	GGTGACTGTG	GATGGCGGTC	GTGGGGTCTG	3150
	ATGTGTGGTG	ACTGTGGATG	GGGTCTGTTG	GGTCTGATGT	GGTGACTGTG	GATGGCGGTC	GTGGGGTCTG	3220
50	GGGTCTGATC	TGTGGTGACT	GTGGATGGCG	GTCTGTTGGT	CTGATGTGGT	GACTGTGGAT	GGCGGTCTGT	3290
	GGGTCTGATC	TGTGGTGACT	GTGGATGGCG	GTCTGTTGGT	CTGATGTGGT	GACTGTGGAT	GGCGGTCTGT	3360
	GTACACAGGG	TCTGATGTGT	GGTGACTGTG	GATGGCGGTC	GTGGGGTCTG	ATGTGTGGTG	ACTGTGGATG	3430
	GCGGTCTGTT	GGTCTGATGT	GTGGTGACTG	TGGATGGCGG	TCGTGGGGTC	TGATGTGTGG	TGACTGTGGA	3500
	TGGCGGTCTG	GGGGTCTGAT	GTGGTGACTG	TGGATGGCGG	TCGTGGGGTC	TGATGTGTGG	TGACTGTGGA	3570
55	GCAGGTGGAG	TCCCAGGTGT	GTCTGTAGCT	ACTTTGCGTC	CTCGGCCCTC	CGGGCCCCCT	TTCCCAAACA	3640
	GAAAGCTTCC	AGGCGCTCTC	TGGGCTTTCAT	CCCGCCATCG	GGCTTGGCCG	CAGGTCCACA	CGTCTGATC	3710
	GGAAGAAACA	AGTGCCACAG	CTGCGCCGGG	GCAGGCCACA	TTTGTGGCTC	ATGCCCTCTC	CTCTGCCGGC	3780
	AGGTCTCTAC	CTTGACAGAC	CTCCAGCCGT	ACATGCCACA	GTTCGTGGCT	CACCTGCAGG	AGACCAGCCC	3850
	GCTGAGGGAT	GCCGTCGTCA	TCGAGCAGGT	CTGGGCACCT	CCCTGCAGGG	TGGGACACGG	ACTCCCAGCA	3920
60	GTGGGTCTCT	CCCTGGGCAA	TCACTGGGCT	CATGACCGGA	CAGACTGTGT	GCCCTGGGGG	GCAGTGGGGG	3990
	GAATGAGCTG	TGATGGGGGC	ATGATGAGCT	GTGTGCTTTG	GCGAAATCTG	AGCTGGGCTG	TGCCAGGCTG	4060
	CGACAGCTGC	TGCATTACAG	CACCTGCTCA	CGTTGACTG	CGCGGCTCTT	CTCCAGTTCC	GCAGTGCCTT	4130
	TGTTTCATGAT	TTGCTAAATG	TCTTCTCTGC	CAGTTTGTAT	CTTGAGGCCA	AAGGAAAGGT	GTCCCCCTCC	4200
	TTTAGGAGGG	CAGGCCATGT	TTGAGCCGTG	TCCTGCCCTG	CTGGCCCCCT	AGTGTGGGTG	CTGAGGCCAA	4270
65	AGGAAACGTG	TCCCCCTTCT	TAGGAGGACG	GGCCGTGTTT	GAGCCACGCC	CCGCTGAGCG	GGCCTCTCAG	4340
	TGCTGGGTCT	GTCCACGTGG	CCCTGTGGCC	CTTTGCAGAT	GTGGTCTGTC	CACGTGGCCC	TGTGGCTCTT	4410
	TGCAGATGCC	TGTTAGCACT	TGCTGGGCTC	TAGGGGACAG	TCGTGTCCAC	CGCATGAGGC	TCAGAGACCT	4480
	CTGGGCGAAT	TTCTTTGGCT	CCAGGGTGGG	GGGTGGAGGT	GGCCTGGGCT	GCTGGGACCC	AGACCTCTGT	4550
	CCCCGCAGCT	GGGCAGCAAC	TCTTGGATCA	CATATGCCAT	CCGGGCCACG	GTGGGCTGTG	TGGGTGTGAG	4620
70	CCCAGCTGGA	CCCACAGGTG	GCCCAGAGGA	GACGTTCTGT	GTACACACTC	CTGCCTAAGC	CCATGTGTGT	4690
	CTGCAGAGAC	TCGGCCCGGC	CAGCCACAGA	TGGCCCTGCA	TTCCAGCCCA	GCCCCGCACT	TCATCACAAA	4760
	CACTGACCCC	AAAAGGGACG	GAGGGCTTTG	GCCACGTGGT	CCTGCCTGTC	TCAGCACCCA	CCGGCTCACT	4830
	CCCATGTGTC	TCCCGTCTGC	TTTCGACAGG	CTCCTCCCTG	AATGAGGCCA	GCAGTGGGCT	CTTCGACGTC	4900
	TTCTTACGCT	TCAATGTGCA	CCACGCCGTG	CGCATCAGGG	GCAAGTAGGT	CAGGTGGCCA	GGTGCCATTG	4970
	CCCTCGGGGT	GGCTGGGCGG	GCTGGCAGGG	CTTCTGCTCA	CCTCTCTCCT	GCCCCCTCCC	CAGTGNCTTT	5040

	CTGCCCGGGG	CCACCAGAGT	CTCCTTTTCT	GGCCCCGGCC	CCCTCCGGCT	CCTGGGCTGC	AGGCTCCCGA	5110
	GGCCCCGGAA	ACATGGCTCG	GCTTGGCGCA	GCCGGAGCGG	AGCAGGTGCC	ACACGAGGCC	TGGAAATGCG	5180
	AAGCCGGGTG	TGGAGTTGCT	CCTGCGTGGA	GGACGAGGGG	CGGGGGGTGT	GTCTGGGTCA	GGTGTGCGCC	5250
	GAGCGTTTGA	GCCTGCAGCT	TGTCAGCTCC	AAGTTACTAC	TGACGCTGGA	CACCCGGCTC	TCACACGCTT	5320
5	GTATCTCTCT	CTCCCCGATC	AAAAGGATTT	TATCCGATTC	TCATTCTGTG	CCCTGTCGTG	TGACCCCCCG	5390
	GAGGGCGCGG	GCTCTTCTCT	CTGTGACTAG	ATTTCCCATC	TGGAAGATGC	GGGGTTGACC	GTGTAGTTTG	5460
	CTCCTCTCTG	GGGGCCTGTG	GTGGCCATGG	GGCAGGCGGC	CTGGGAGAGC	TGCCGTGACA	CAGCCACTGG	5530
	GTGAGCCACA	CTCACGGTGG	TAGAGCCACA	GTGGCTGGTG	CCACATCAGC	TCTCTCTGGT	TTTAAAGTAA	5600
	ACCACACACC	TCCCGGCGAG	CATCTGCCTG	CGACCCGTGT	TGTGCTGGG	GAGAGTGGTA	GCACGGAGGA	5670
10	AATTCGTGCA	CACCTCAAGT	CATCAGCAAG	GTCAATCCGA	GTCAAGTGGG	ACGTGGAGGC	CTCTCTCTGG	5740
	GATCGTCTCC	AGCGGATAAA	GGACTGTGCA	CAGCTTCGGA	AGCTTTTATT	TAAAAATATA	ACTATTAAAT	5810
	ATTGCAATTAT	AAGTAATCAC	TAATGGTATC	AGCAATTATA	ATATTTATTA	AAGTATAATT	AGAAATATTA	5880
	AGTAGTACAC	ACGTTCTGGA	AAAACACAAA	TTGCACATGG	CAGCAGAGTG	AATTTTGGCC	GAGGGACACG	5950
	TGTGCACATG	TGTGTAAGCG	GCCCCAGGCC	CCACAGAAAT	CGCTGACAAA	GTCACTCTCC	CAGAGAAGCC	6020
15	AGTCAGGGCC	TCTTTCGTGG	TCGTGAATTT	TATTAAGATG	GATCAAGTCA	CGTACCGTCC	ACGTGTGGCA	6090
	GGGCTTTGGG	GAATGTGAGG	TGATGACTGC	GTCTCTATGC	CCTGACAGAC	AGGAGGTGAC	TGTGTCTGTC	6160
	CTGTCCCTAG	GACACGGACA	GGCCCGAAGC	TCTAGTCCCC	ATCGTGGTCC	AGTTTGGCCT	GTGAATAAAA	6230
	ACGTCTTCAA	AACCTGTTGC	CCCAAAAACT	AAGAACAGAG	AGAGTTTCCC	ATCCCATGTG	CTCACAGGGG	6300
20	CGTATCTGCT	TGCGTTGACT	CGCTGGGCTG	GCCCGACTCC	TAGAGTTGGT	GCGTGTGCTT	CTGTGCAAAA	6370
	AGTGCACTCC	TCTTGGCCAT	CACGTGTGATA	TCTGCACCAG	CAAGGAAAGC	CTCTTTTCTT	TTCTTTCTTT	6440
	TTTTTTTTTT	GAGACGGAAC	GTCACTGTGG	TCTGCCTGGG	CTTGAGTGCA	GTGGCGCGAT	CTCAACTCAC	6510
	TGCAACCTCC	GCCTCCCGGG	TTCAGCAATT	TCTCTGCTCT	CAGCCTCCCG	AGCAGCTGAG	ATTACAGGCA	6580
	CCCAACCCCT	GCGCCTGGCT	AATTTTGTGA	TTTTTAGTAG	AGAGGGGTTT	TTGCCATGTT	GGCCAGGCTG	6650
25	GTCTCGAACT	CCTGACCTCA	GGTGATCCAC	CCACCTCGGC	CTCCCAAAGT	GCTGGGATTA	CAGGTGTGAG	6720
	CCATCAGGCC	CAGCCGGAAG	GCCTCTTTTT	AAGGTGACCA	CCTATAGCGC	TTCGCCAAAA	TAACAGGTCT	6790
	TGTTTTTGCA	GTAGGCTGCA	AGCGTCTCTT	AGCAACAGGA	GTGGCGTCTC	GTGGGCTCTG	GGGATGGCTG	6860
	AGGGTCGCGT	GGCAGCCATG	CCTTCTGTGT	GCACCTTTAG	GTTCACAGGG	GCTATTCTGC	TCTCACTGTT	6930
	TGTCTGAAAA	CGCACCCCTG	GCATCCTTGT	TTGGAGAGTT	TCTGCTTCTC	GTTGGTCATG	CTGAAACTAG	7000
30	GGGCAAGGTT	GTATCCGTTG	GCGGCGAGCG	GCTACATGTA	GGGTGATGAG	TCTTTCACCG	TGGCAAAATT	7070
	CCTTGAAAAA	AAAAAAAGGA	GTCCGGTTAA	GCATTCAITC	CGGGTCAAGT	GTCTGTTTCT	GTGAATAAAC	7140
	TCTAAGATT	AAGAAACCTT	AATGAAGAA	AACCTTGATG	ATTACAGACA	AGGATGTGGT	CACACCTGTG	7210
	GCTGGATCTG	TTTCAGCCGC	CCCAGTGCAT	GGTGAGAGTG	GGGAGCAGGG	ATTGTTTGT	CAGAGGTCTC	7280
	ATCTGGTATG	TTTCTGAGGT	GTTCGCCGCG	TGAATGGTAG	ACGTGTCGTT	TGTGTGTATG	AGGTTCTGTG	7350
35	TCTGTGTGTG	GCTCGGTTTG	AGTGTACGCA	TGTCCAGCAC	ATGCCCTGCC	CGTCTCTCAC	CTGTGCTTTC	7420
	CCGCCCCAGG	TCCTACGTCC	AGTGCCAGGG	GATCCCGCAG	GGCTCCATCC	TCTCCACGCT	GCTCTGCAGC	7490
	CTGTGCTACG	GCGACATGGA	GAACAAGCTG	TTTGGCGGGA	TTGGCGGGGA	CGGGTGAGGC	CTCTCTTCC	7560
	CCAGGGGGGG	TTGGGTGGGG	GTGATTTTGC	TTTTGATGCA	TTCACTGTGA	ATATTCCTGG	TGCTCTGGAG	7630
	ACCATGACTG	CTCTGTCTTG	AGGAACACGA	CAAGGTTGCA	GCCCTTCTTT	GGTATGAAGC	CGCACGGGAG	7700
40	GGGTTCACCA	GCCTGAGGAG	TGCGGGCTCC	ACGCAGGCTC	TGTCACGCGG	CCATGTCCAG	AGGCTCTCAG	7770
	GCTCAGCAGG	CGGGAGGGCC	GCTGCCCTGC	ATGATGAGCA	TGTGAATTCA	ACACCGAGGA	AGCAACACAG	7840
	CTTCTGTAC	GTACCCAGG	TTCCGTTAGG	GTCCCTGGGG	AGATGGGGCT	GGTGACGCTC	GAGGCCCCAC	7910
	ATCTCCACGC	AGGCCCTCGA	CAGGTGGCCT	GGACTGGGCG	CCTCTTCAGC	CCATTGCCCA	TCCCACTTGC	7980
	ATGGGGTCTA	CACCCAAGGA	CGCACACACC	TAAATATCGT	GCCAACCTAA	TGTGGTTCAA	CTCAGCTGGC	8050
45	TTTTATTGAC	AGCAGTTACT	TTTTTTTTTT	TAATACTTTA	AGTTCTAGGG	TACATGTGCA	CGACGTGCAG	8120
	GTTAGTTTACA	TATGTATACA	TGTGCCATGT	TGGTGTGCTG	CACCCATTAA	CTCATCATTT	ACATTAGGTA	8190
	TATCTCTCTAA	TGCTATCCCT	CCCCACTCCC	CCCATCCCAT	GACAGGCCCT	GGTGTGTGAT	GTTCGCCACC	8260
	CTGTGTCCAA	GTGTTCTCAT	TGTTCACTTC	CCACCTGTGA	GTGAGAACAT	GTGGTGTGTT	GTGTTCTTTC	8330
50	CTTGCAATAG	TTTGCTCAGA	GTGATGGTTT	CCAGCTTCGT	CCATGTCCCT	ACAAAGGACA	GTAACCTATC	8400
	CTTTTTTATG	ACTGCATAGT	ATTCCGTGGT	GTATATGTGC	CACATTTTCT	TAATCCAGTC	TATCATCGAT	8470
	GGACATTG	GTTGGTTGCA	AGTCTTTGCT	ACTGTGAATA	GTGCCCAAT	AAACATACGT	GTGCATGTGT	8540
	CTTTATAGCA	GCATGATTTA	TAATCCTTTG	GGTATATACC	CAGTAATGGG	ATGGCTGGGT	CAATGGTGTG	8610
	TTCTAGTTCT	AGATCCTTGA	GGAATCACCA	CACGTCTCTC	CACAATGGTT	GAACATAGTT	ACACTCCAC	8680
55	CAACAGTGTA	AAAGTGTTC	GGTGCTGGAG	AGGATGTGGA	CAGCAGTTAT	TTTTTTATGA	AAATAGTATC	8750
	ACTGAACAAG	CAGACAGTTA	GTGAAGGATG	CGTCAGGAAG	CCTGCAGGCC	ACACAGCCAT	TTCTCTCGAA	8820
	GACTCCGGGT	TTTTCTGTG	CATCTTTTGA	AACTCTAGCT	CCAATTATAG	CATGTACAGT	GGATCAAGGT	8890
	TCTTCTTCAT	TAAGGTTCAA	GTTCTAGATT	GAATAAGTT	TATGTAACAG	AAACAAAAAT	TTCTTGATCA	8960
	CACAACCTGC	TCTGGGATTT	GGAGGAAAGT	GTCTCGAGC	TGGCGGCACA	CTGGTCAGCC	CTCTGGGACA	9030
	GGATACCTCT	GGCCCATGGT	CATGGGGCGC	TGGGCTTGGG	CCTGAGGGTC	ACACAGTGCA	CCATGCCCGAG	9100
60	CTTCTGTGG	ATAGGATCTG	GGTCTCGGAT	CATGCTGAGG	ACCAAGAGTG	CCATGCTGGT	AAAGGGCACC	9170
	ACGTGGCTCA	GAGGGGGCGA	GGTCCGAGC	CCAGCTTTTC	TTACCGTCTT	CAGTTATTTT	TCCTTAAGAG	9240
	TCTGAGAAAT	GGGGCCGCGC	CTGATGGCCT	TCGTTCTGCT	TCAGCTGGCA	CAGAATTGCA	CAAGCTGATG	9310
	GTAAACACTG	AGTACTTATA	ATGAATGAGG	AATTGCTGTA	GCAGTTAACT	GTAGAGAGCT	CGTCTGTTGG	9380
	AAAGAAATTT	AAGTTTTC	TTTAAACCGT	TTGGAGAATG	TTACTTTTAT	TATGGCTGTG	TAAATGTTT	9450
65	GACATTCAGT	CCCTCGTAGA	CAGATACTAC	GTAAAAAGTG	TAAAGTTAAC	CTTGCTGTGT	ATTTTCCCTT	9520
	ATTTTAGGCT	GCTCCTGCGT	TTGGTGGATG	ATTCTTGT	GGTGACACCT	CACCTCACCC	ACGGGAAAAAC	9590
	CTTCTCAGG	TGAGGCGCGT	GCGTGTGTCT	TGTGGGGACC	TCCACAGCCT	GTGGGCTTTG	CAGTTGAGCC	9660
	CCCCGTGTCC	TGCCCTGGC	ACCGCAGCGT	TGCTCTGCTC	AAGTCTCTC	TCTCTGCGGG	TGCTGGATCC	9730
	GCAAGAGCAG	AGGCGCTTGG	CCGTGCACCC	AGGCCCTGGG	GCGCAGGGGC	ACCTTCGGGA	GGGAGTGGGT	9800
70	ACCGTGCAAG	CCCTGTGCTC	GCAGAGACGC	ACCCAGGTTA	CACACGTGGT	GAGTGACAGC	GGTGACCTGG	9870
	CTCCTGCTGC	TCTTTGGAAA	GTCAAGAGTG	GCGGCTCCTG	GGGCCCCAGT	GAGACCCCCA	GGAGCTGTGC	9940
	ACAGGGCCTG	CAGGGCCGAG	GCGGCAGCCT	CCTCCCCAGG	GTGCACCTGA	GCCTGCGGAG	AGCAGGAGCT	10010
	GCTGAGTGAG	CTGGCCACAC	GCGTTCGCTG	CGGTACAGTT	CCTGCGTGGG	GTTGTTTGGG	ATCGGTGGGA	10080
	GAATTTGGAT	TGCTGAGTGT	CTGCTGTCTT	GAACACCGGA	GATGGCTAGG	AGTGGGTTTC	AGAGTTGATT	10150
75	TTTGTAATC	AAACTAAAT	CAGGCACAGG	GGACCTGGCC	TCAGCACAGG	GGATTGTCCA	ATGTGGTCCC	10220
	CCTCAAGGGC	GCCCCACAGA	GCCGGTGGGC	TGTTTTTAAA	GTGCGATTG	ACGAGGGAGC	AGAAACCTTG	10290
	AAAGCTGTAA	AGGGAACCTT	CAGAAAATGT	GGCCGCCAGG	GGTGGTTTCA	GGTGCTTTGC	TGGGCTGTGT	10360
	TTTGAAAAAC	CCATTGGAC	CCGCCCTCCA	AGTCCACCTC	CCAGGTCCAC	CCTCCAGGGC	CGCCCTGGGC	10430

	TGGGGGTATG	CCTGGCGTTC	CTTGTGCCGC	AGCCCGGAGC	ACAGCAGGCT	GTGCACATT	AAATCCACTA	10500
	AGATTCACCTC	GGGGGGAGCC	CAGGTCCCAA	GCAACTGAGG	GCTCAGGAGT	CCTGAGGGCTG	CTGAGGGGAC	10570
	AGAGCAGACG	GGGAACCGTG	CTTCTGTGTG	GCAAGTTCCT	GAGGGTGTCTG	GCCAGGGAGG	TGGCTCAGAG	10640
5	TGTATGTTGG	GGTCCCACCG	GGGGCAGAAC	TCTGTCTCTG	ATGAGTCGGC	AGCCATGTAA	CAGGAAGGGG	10710
	TGGCCACAGG	GAGCTGGGAA	TGCACCCAGG	GAGCTGCCGA	GCTGGCCGAG	GTCCCAGGGC	CAGGCCACAG	10780
	GAAGGGCAGG	GGGACGCCCG	GGGCCACAGC	AGAGGCCGCA	GGAAAGGGAAG	GGGATGCCCA	GGCCAGAGCA	10850
	GAGGCTACCG	GGCACAGGGG	GGCTCCCTGA	GCTGGGTGAG	CGAGGCTCAT	GACTCGGCGA	GGCAACCTCC	10920
	TTGACGTGAA	GCTGACGACT	GGTGTGCCCC	AGCTCACAGC	CCAGCCAGGT	CCCGCGCCTG	AGCAGGAATC	10990
10	CAGAACCCTC	CCCTTTGTCT	AAAGCACAGC	AGATGCCTTC	AGGGCATCTA	GGAGAAAACA	GGCAAAAGTCG	11060
	TTGAGAAACG	TCTTAAAGA	AGGTGGGATG	GTGGCAATTT	CTTGTCCAGA	TTTATGTCCTG	CCCCGGACCA	11130
	CAGATGAGTC	TATAACGGGA	TTGTGGTGT	GCCATGGGGA	CACATGAGAT	GGACCATCAC	AGAGGCCACT	11200
	GGGGCTGCAC	CTCCCATCTG	AGTCCTGGCT	GTCCCGGGTC	CAGGCCAGGT	TCTTGATGTC	TCACCTACCT	11270
	GTCTGCCCG	GGAGACAGGG	AAAGCACCCC	GAAGTCTGGA	GCAGGGCTGG	GTCCAGGGTC	CTCAGAGCTC	11340
15	CTGCCAGGCC	CAGCACCTCT	CTCCAAATCA	CCACTTCTCT	GGGGTTTTC	AAAGCATTTA	ACAAGGGTGT	11410
	CAGGTTACCT	CCTGGGTGAC	GGCCCCGCAT	CTGGGGCTG	ACATTGCCCC	TCTGCCTTAG	GACCCTGGTC	11480
	CGAGGTGTCC	CTGAGTATGG	CTGCGTGGTG	AACTTGCCGA	AGACAGTGGT	GAACCTCCCT	GTAGAAGACG	11550
	AGGCCCTGGG	TGGCACGGCT	TTTGTTCAGA	TGCCGGCCCA	CGGCCTATTG	CCCTGGTGGC	GGCTGTGCT	11620
	GGATACCCCG	ACCTTGGAGG	TGCAGAGCGA	CTACTCCAGG	TGAGCGCACC	TGGCCGGAAG	TGGAGCCTGT	11690
20	GCCCCGCTGG	GGCAGGTGCT	GCTGCAGGGC	CGTTGCGTCC	ACCTCTGCTT	CCGTGTGGGG	CAGCGGCTGT	11760
	CCAATCCCAA	AGGGTCAGAG	GCCACAGGGT	GCCCCCTGCT	CCATCTGGGG	CTGAGCAGAA	ATGCATCTTT	11830
	CTGTGGGAGT	GAGGTGCTC	ACAACGGGAG	CAGTTTTCTG	TGCTATTTTG	GTAAGAGGAA	ATGGTGCACC	11900
	AGACCTGGGT	GCACTGAGGT	GTCTTCAGAA	AGCAGTCTGG	ATCCGAACCC	AAGACGCCCG	GGCCCTGCTG	11970
	GGCGTGAGTC	TCTCAAACCC	GAACACAGGG	GGCCTGTGG	GCATGAGTCC	CTCTGAACCC	GAGACCTCTG	12040
25	GGCCCTGCTG	GGCGTGAGTC	TCTCCGAACC	CAGAGACTTC	AGGGCCCTTT	TGGGCGTGAG	TCTCTCCGCT	12110
	GTAGGCCCCA	CAGTCCAAGG	CTCATCCACA	GTCTACAGGA	TGCCATGAGT	TCATGATCAC	GTGTGACCCA	12180
	TCAGGGGACA	GGGCCATGGT	GTGGGGGGGG	TCTCTACAAA	ATTCTGGGGT	CTTGTTTCCC	CAGAGCCCCA	12250
	GAGCTCAAGG	CCCCGTCTCA	GGCTCAGACA	CAAATGAATT	GAAGATGGAC	ACAGATGCAG	AAATCTGTGC	12320
30	TGTTTCTTTT	ATGAATAAAA	AGTATCAACA	TTCCAGGCAG	GGCAAGGTGG	CTCACACCTA	TAATCCCAGC	12390
	ACTTTGGGAG	GCCGAGGTGG	GTGGATCACT	TGAGGCCAGG	AGTTTGAGGC	CAACCTAACC	AACATAGTGA	12460
	AATTCATT	CTACTTAAAA	AATACAAAAA	TTAGCTTGGC	CTGGTGGCAC	ACGCCTGTAG	TCCCGCTAT	12530
	GCGGGAGGCT	GAGGCAGGAG	AATCATTTGA	ACCCAGGAGG	CAGAGGTTGC	AGTGAGCCGA	GATCACACCA	12600
	CTGCATCTCA	GGCTGGGCAA	CAGAGTGAGA	CTTCATCTTA	AAAAAAGGAA	AAAAAGTATC	AGCATTTCCA	12670
35	AACCATAGTG	GACAGGTGTT	TTTTTATTTT	GTCCCTCGAT	AATATTTACT	GGTGCTGTGC	TAGAGGCCGG	12740
	AACCTGGGGT	GCCTTCCTCT	GAAAGGCACA	CCTTCATGGG	AAGAGAAATA	AGTGGTGAAT	GGTTGTTAAA	12810
	CCAGAGGTTT	AAACTGGGGT	CCTGTCTGTC	TGAGTTAACA	GTCCAGATCT	GGACTTTGCC	TCTTTCCAGA	12880
	ATGCTCCCTG	GGGTTTGCTT	CATGGGGGAG	CAGCAGGTGT	GGACACCCCT	GTGATGGGGG	AGCAGCAGGT	12950
	GCAGACGCCC	TCATGATGGG	GGAGTGGCAG	GTGCAGACAC	CCTTGTGCAT	GGTGCCGAGC	ATGTCCCTGT	13020
40	TGCAGCTCCC	TCCCAACAAG	GATGCCGGTC	TCTGTGCTC	CCACAGTCC	CTGCTTCCCT	CTCAGCCCT	13090
	TACCTGGTCC	TGGCCTCCAC	TGGCTTTGTC	TGCATGATT	CCACATTCCC	TGGGCTCCCA	GCACCTCTTC	13160
	GCCTCTCCCA	GGCACCTCTG	CAGTGCTGGC	CATACCAGTC	AGCTGTGAAC	TGCTCACTGC	TTATTTTGCT	13230
	CCCCATGAAA	TGTATTTTTT	AGGACAGGCA	CCCTGGTTTC	CAGCCTCTGC	CACAGCATCA	GTGAATGTTA	13300
	TTGAAGGACA	AAGGACAGAG	AAACAAATCA	GGAAATGGG	TTCTCTCTAA	ACACATTGCA	AAGCCACAGA	13370
45	GGCTAGTGCA	GGATGGGGTG	GCATCAGGTC	ATCAGATGTG	GGTCCAATGC	CAGAATATTC	TGTGCTCCCA	13440
	AAGGCCACTT	GGTCAGAGTG	TGTGCTTGCA	GAGGTGGCTC	TAAAGGCTCA	GCAGTGGAGG	CAGTGGTTCG	13510
	CCATACTCAG	GGTGAACCTA	CATCCTCTGC	GTCTGAAGTA	TACAGCAGAG	GCTTGAAGGG	CACTGTGGAG	13580
	AGAAAAACAG	GCAAAATGAT	TAAAGAAAGT	GAAAAAGGAA	AAGTGGTAA	ATGGGAATTT	TCTTGTCCAG	13650
	ATTTTAGTCT	CCCCAACAC	AGCTCAGATG	GTAGAATGTG	GTCAAGAACTG	ATGGACAGAA	CAATAGAACA	13720
50	AAACGGGAAGC	CCTATCTCTC	AGAAACGTGT	GTTAATGTGG	TATGTGGCAC	AGCTGATGGA	AAAGAGAGTG	13790
	TGTGTGAAT	TTTTTTTTCT	GAGAAAACTG	ACTGGAAGCA	AATAAGTTGT	GTCTTTACAG	CATATACCAG	13860
	AGCAGATTCT	AGGTAGAAGA	GGAGACACAT	GCAAAACAAC	CCAGCAACAG	AAATAAACA	AAAGACTCAA	13930
	AGGGAAGGGA	GGTGAACGTT	CCCTGGTTTG	CTGTTGGGGA	AGGACACACA	GGGAGGCCGA	TGAAACCAAGT	14000
	GAGGCAACGG	GCATTGCTTT	CAGTGCAGAG	AAACTCAGCT	TGCCCTGAGC	ACAGTGAAAA	TGGCCATTCC	14070
55	CTGGAGCGTT	TGTGCACGTG	ATTTATTTAA	GGCGCCCTGT	GAGGTCTCTG	ACATTATCCC	TCTCACTTTG	14140
	TTCTCCTAAC	CACCTGAGAG	GTAGAGGAGG	AAAGGCTCCA	GGGGAGCAGC	CGCCCTTGGT	CACCCAGCTG	14210
	GCAAAAGGCA	TGCATGATTG	CAGCCTGGCC	TCCTGCTCCG	GGGCCCTTGC	TCTGCCCGAG	GACCCACAC	14280
	AAGTCAGACC	CATAGGCTCA	GGGTGAGCCG	GAGCCCAAGG	TCGTGTTGGG	GATGGCTGTG	AAAGAGAA	14350
	TGGACGCTCG	ATGCACACTT	GGGAAGGTCC	TACCAGCAGC	GTCAAAGAAA	TGCATGTGAA	ACTGACAGCG	14420
60	AGACCCATCC	CTCAAGAA	CGCACGTGAA	ACTGATGGCG	AGACCTGTCC	CCATCCCTCA	TGCTGGCTCC	14490
	TTTTCTGGGC	TTGCCAAGAG	CCAGCATCAG	GTGAGGCA	GCTGGAAGA	CTTTTCTGGA	AAGCAGCTTG	14560
	TTTGATGGA	AGTCCTCACA	ATGTCTGTG	TCTTCCAGT	AATTCCACTT	CTGAAGTGAC	CAGACATTAT	14630
	CACGGGTCTT	ATTACCAT	TCCAGTGTTC	CAGGCAGGGG	GACTTGCCAC	AGCAAGTCAC	GAACCTGCCC	14700
65	AAATACAGGG	CTAAGGAGAT	ATTATGCATC	ACAAAACCTG	CTCTGCCATT	AAACATTTTT	CAAAGAAATTT	14770
	TTGAAGAATG	TTTAATGGCA	CAAAACGTTT	ATTCAATGT	AGCAGTGTTC	AAAGCTGGAT	GTAAGAAGAAC	14840
	ACACCCACAGG	AGCCTGCGGT	GAATGTCTATG	TGTGTTCTATC	TTTGACATG	GACATACATG	GGCAGTGAGT	14910
	GGTGGTGAGG	CCCTGGAGGA	CATCGGTGGG	ATGCCCTCAT	CCTGCCCTC	TGGAGACACC	ATGTGTGCCA	14980
70	CGTGCACTCA	CTGGAGCCCT	GTTTAGCTGG	TGCCACCTGG	CTCTTCCATC	CCTGAGATTG	AAACACAGTG	15050
	AGATTCCCCA	CGCCCAACTC	AGTGTCTCTC	CACAAAAAAC	CTGAGTCACA	CCTGTGTTCA	CTCAGGGGAC	15120
	GCCCCGGAGC	CAGGGCTCCA	CAGTTTATTA	TGTGTTTTTG	GCTGAGTTAT	GTGCAGATCT	CATCAGGGCA	15190
	GATGATGAGT	GCACAAACAC	GGCCGTGCGA	GGTTTGATA	CACCTCAACAT	CACCTAGCCAG	GTCTGGTGG	15260
	AGTTTGGTCA	TGCAGAGTCT	GGATGGCATG	TAGCATTGG	AGTCCATGGA	GTGAGCACCC	AGCCCTGCTG	15330
75	GGCTGCAGCG	CATGCCCCAG	GCAGGACAAG	GAAGCGGGAG	GAAGGCAGGA	GGCTCTTTGG	AGCAAGCTTT	15400
	GCAGGAGGGG	GCTGGGTGTG	GCGCAGGCAC	CTGTGTCTGA	CATTCCCCC	TGTGTCTCAG	ATCTGCCCGG	15470
	ACCTCCATCA	GAGCCAGTCT	CACCTTCAAC	CGCGGCTTCA	AGGCTGGGAG	GAACATGCGT	CGCAAACTCT	15540
	TTGGGGTCTT	GCGGCTGAAG	TGTCAACAGC	TGTTTCTGGA	TTTGCAAGTG	AGCAGGCTGA	TGTGACGAC	15610
	AGAGTTTACA	GTTTCAAGAGG	TGTGTGCGCA	AGTATGTGTG	TGTGTGTGTG	CGCGGCTGCC	TGCAAGGCTG	15680
	ATGGTGACTG	GCTGCACGTA	AGAGTGCACA	TGTACGCATA	TACAGCTGAG	CACATACATG	TGTGCATGTG	15750
	TGTACATGAA	GGCATGGCAG	TGTGTGCACA	GCTGTGCAAG	GGCACAAGTG	TGTGCACATG	CGAATGCACA	15820

5

10

15

20

25

30

35

40

45

50

55

60

65

70

75

CCTGACATGC ATGTGTGTTT GTGCACAGTC GTGTGGGCAT TCACGTGAGG TGCAATGCGTG TGGGTGTGCA 15890
 GTGTGAGTAG CATGTGTGCA CATAACATGT ATTGAGGGGT CCTCGTGTTC ACCCGCGTAG GTCCTCAGCA 15960
 CCAGTGCCAC TCCTTACAGG ATGAGACGGG GTCCACAGGC TTGGTGGGCT GAGGCTCTGA AGCTGCAGCC 16030
 CTGAGGGCAT TGTCCCATCT GGGCATCCGC GTCCACTCCC TCTCTGTGG GCTTCTGTGT CCACTCCCCC 16100
 TCTCCTGTGG GCATTATCAT CCACTCCACT CCTCTCTCC TGTGGGCATC CGCGTCCACT CCCCTCTCT 16170
 GTGGGCATCT GCGTCCACCT CCCCTCTCTG TGGGCATTG CGTCCACTCC CTCTCTGTGT TCCTTCTGT 16240
 CTTGGCCGAG CCTCGGGGGC AGGCAGATGA CACAGAGTCT TGACTCGCCC AGGGTGGTTC GCAGCTGCCG 16310
 GGTGAGGGCC AGGCCGGATT TCACTGGGAA GAGGGATAGT TTCTTGTCAA AATGTCTCTC TTCTTGTGTC 16380
 CATCTGAATG GATGATAAAG CAAAAAGTAA AAACTTAAAA TCCCAGAGAG GTTCTACCG TTCTCTACTC 16450
 TTTCTTGGCG ACTCTAGGTG AACAGCCTCC AGACGGTGTG CACCAACATC TACAAGATCC TCCTGCTGCA 16520
 GGGCTACAGG TGAGCCGCCA CCAAGGGGTG CAGGCCACGC CTCCAGGGAC CCTCCGGGCT CTGCTCACCT 16590
 CTGACCCGGG GCTTCACTT GGAACCTCTG GGTTTTAGGG GCAAGGAATG TCTTACGTTT TCAGTGGTGC 16660
 TGCTGCCGTG GCACAGTTCT GTTCGCGTGG CTCTGTGCAA AGCACCTGTT CTCCATCTCT GGGTAGTGGT 16730
 AGGAGCCGGT GTGGCCCGAG GTGTCCCCAC TGTGCTGTG CACTGGCCGT GGGACGTCA TGGAGCCATC 16800
 CCAGGGCAGC AGGGGCATGG GGTAAAGAGA TGTTTATGGG GAGTCTTAGC AGAGGAGGCT GGGAAAGGCT 16870
 CTGAACAGTA GATGGGAGAT CAGATGCCCG GAGGATTGGG GGTCTCAGCA AAGAGGGGCG AGGTGGGTGC 16940
 AGGTGAGGGT CGCTGGCCCC ACCCCCGGGA AGGTGCAGCA GAGCTGTGGC TCCCACACA GCCCGGCCAG 17010
 CACCTGTGCT CTGGGCATGG CTGTGCTCTT GGAACGTTCC CTGCTCTGGC TGGTCAGGGG GTGCCCTGTC 17080
 CAAGAATCGA CAACTTTATC ACAGAGGGA GGGCCAATCT GTGGAGGCCA CAGGGCCAGC TTTGTGCTGG 17150
 AGTCAGGGCA GGTGTGGCA CAAGCCTCGG GGCTGTACCA AAGGGCAGTC GGGCACCACA GGCCCGGGCC 17220
 TCCACCTCAA CAGGCCTCCC GAGCCACTGG GAGCTGAATG CCAGGAGGCC GAAGCCCTCG CCCATGAGG 17290
 GCTGAGAAGG AGTGTGAGCA TTTGTGTTAC CCAGGGCCGA GGCTGCCGA ATTACCGTGC ACACCTGATG 17360
 TGAAATGAGG TCGTGTCTA TCGTGGAAAC CCAGCAAGGG CTCACGGGAG AGTTTTCCAT TACAAGGTGC 17430
 TACCATGAAA ATGGTTTTTA ACCCGAGTGC TTGCGCTTC ATGCTCTGGC AGGGAGGGCA GAGCCACAGC 17500
 TGCAATGTTAC CGCCTTTGCA CAGCTCCAG AGGCTTGGGA CCAGGCTGTC TCAGTTCAG GGTGCTCGC 17570
 GCTCAGACCG CCTCTCTCTC TGGCTTCTCT CTCTGCTCA AATCTTCCCT CGTTTGCATC TCCTGACGC 17640
 GTGCTGGGC CCTCTGCAA GCTGCTGAC TCTTTCCCG AAACCTTGG GGTGTGCTGG ATACAGGTGC 17710
 CACTGAGGAC TGGAGGTGTC TGACACTGTG GTTGACCCCA GGGTCCAGCT GGGCTGCTTG GGGCTCTCT 17780
 GGGCCATGAT GAGGTGAGAG GAGTTTTCCC AGGTGAAAC TCTGGGAAA CTCCAGGGC CATGTGACCT 17850
 GCCACCTGCT CCTCCCATAT TCAGCTCAGT CTGTGCTCA TTTCCCCACC AGGGTCTCTA GCTCCAGGA 17920
 GCTCCCGTAG AGGGCCTGGG CTCAGGGCAG GCGGCTGAG TTTCCCCACC CATGTGGGGA CCTTTGGGTA 17990
 GTGCTGTGAT TGGGTAGCCC TGAGGAGGCC GAGATGCGAT GGGCCACGGG CCGTTTCAA ACACAGATGC 18060
 AGGCACGTGG AAGGCCCAGG AATCCCTTC CCTCGAGGCA GGAGTGGGAG AACGGAGAGC TGGGCCCCGA 18130
 TTTACGGGCA GCCAGGCTGC AGTGGGCGAG GCTGTGTGG TCCACGTGGC GCTGGGGGCG GGGTCTGATT 18200
 CAAATCCGCT GGGGCTCGGC CTTCCTGGCC CGTGTGCCC GCGCTCCAC ACGGGCTTGG GGTGGACGCC 18270
 CCGACCTCTA GCAGGTGGCT ATTTCTCCCT TTGGAAGAGA GCCCTCACC CATGCTAGGT GTTCCCTCC 18340
 TGGGTGAGGA GCGTGGCCGT GTGGCAACCC CGGGACCTTA GGCTTATTTA TTTGTTTAAA AACATTCTGG 18410
 GCCTGGCTTC CGTTGTGCT AAATGGGGA AAGACATCCC ACCTCAGCAG AGTTACTGAG AGGCTGAAAC 18480
 CGGGGTGCTG GCTTGACTGG TGTGATCTCA GGTCAATCCA GAAGTGCTC AGGAAGTCAG TGAGACAGG 18550
 TACATGGGGG GCTCAGGCG AGTGGTGAAT GAGGTACACG GGGGGCTCAG GCAGTGGGTG AGGCCAGGTA 18620
 CATATGGGGG TCAGGCACTG GGTGAGATGA GGTACACGGG GGGCTCAGGC AGAGGGTCAG ACCAGTACA 18690
 CGGGGGCTCT GATCACACGC ACATATGAGC ACATGTGCAC ATGTGCTGTT TCATGTAGC CAGGTCTGTG 18760
 CACACCTGCC CCAAGTCCCG AGGAAGCTGA GAGGCCAAG ATGGAGGCTG ACAGGGCTGG CCGCTGGGT 18830
 CACACCTGTA GTCCAGCAC TTTGGGAGGC CGAGGCGAGA GGATCCCTTG AGCCACAGGAG TTTAAGACCA 18900
 GGCTGAGCAA CATAGTAGAA CCCCATCTCT ATGAAAAATA AAAACAAAAA TTAGCTGAAC ATGGTGGTGT 18970
 GCGCTGTAG TTCCARTACT TGGGAGGCTG AAGTGGGAGG ATCACTTGAG CCCAGGAGGT GGAAGTGC 19040
 GTGAGCTGAG ATTGCACAC TGTACTGCAG CCTGGGTGAC AGAGTGAGAG CCCATCTCAA CAACACAAA 19110
 GAAGACTGAC AAATGCAATT TCTTGAAGAG AAACATTAG TAGGAACCTA ACCTACACAC AGAAGCCAAG 19180
 TCGGTGTCTC GGTGTCACTG AGATGAGATG ATGGGTCTCT ACACCATCAC CCCAGACCA GGGTTATGTC 19250
 ACCACAGGGG CCGGTGGCTC AGAAGGATG CCGAGGACGT TGATATACGA TGACATCAAG GTTGTCTGAC 19320
 GAAGGGCAGG ATTCATGATA AGTACCTGCT GGTACACAA GAAACATGGA TAAACTGGA ACCTTAGAGG 19390
 CCTTCCCGGA ACAGGGGCTA ATCAGAAGCC AGCATGGGGG GCTGGCATCC AGGATGGAGC TGCTTACGCC 19460
 TCCACATGCG TGTTCATACA GATGGTGCAC AGAAACGCAG TGTACTGTG CACACACAGA CACGACGTA 19530
 CTCGCACACA CAAGCACACA CACAGACATG CATGATGCA TCCGTGTGTG TGCACCTGTG CCCATGAGGA 19600
 AACCCATGCA TGTGCATTCA TGCACGCACA CAGGCCACGG TGGGCCCATG CCCACACCA CGAGCACCGT 19670
 CTGATTAGGA GGCTTTCTCT CTGACGCTGT CCGCATCTCT CTAGGTTTC ACGCATGTGT GCTGCAGCTC 19740
 CCATTTCTAT AGCAAGTTTG GAAGAACCCC ACATTTTTC TCGCGGTGAT CTCTGACAGC GCCTCCCTCT 19810
 GCTACTCCAT CCTGAAAGCC AAGAAGCGAG GTATGTGAG GTGCTTGGCC TCAGTGGCAG CAGTGCCTGC 19880
 CTGCTGGTGT TAGTGTGTCA GGAGACTGAG TGAATCTGGG CTTAGGAAGT TCTTACCCCT TTTGCTATCA 19950
 GGAAGTGGTT TAACCCAACC ACTGTACGGC TCGTCTGCCC GCCCTCTCTG GGGGTGAGCA GAGCACCTGA 20020
 TGGAGGGGAC AGGAGCTGTC TGGGAGCTGC CATCTTCCC ACCTTGTCTG GCCTGGGGAA GCGCTGGGG 20090
 GCCTGGTCTC TCTGTTTTC CCTATGGTGG GATTTGGGGG GCCTGGCCTC TCCTGTTTGC CCGTGGTGG 20160
 GATTGGGCTG TCTCCCGTGC ATGGCACTTA GGGCCCTGT GCAAAACCCAG GCCAAGGGCT TAGGAGGAGG 20230
 CCAGGCCCAG GCTACCCAC CCCTCTCAGG AGCAGAGGCC GCGTATCAC ACACAGAGC CCCGCGCGCT 20300
 CCTCTGCTTC CCAGTACCCG TCCTCTGCCC TGGACACTT TGTCCAGCAT CAGGGAGGTT TCTGATCCGT 20370
 CTGAAATTTA AGCCATGTGC AACCTGCGGT CCTGAGCTTA ACAGCTTCTA CTTTCTGTTC TTTCTGTGTT 20440
 GTGGAATTT CACCTGGAGA AGCCGAAGAA AACATTCTG TCGTACTCC TGCGGTGCTT GGGTGGGAG 20510
 AGCCAGAGAT GGAGCCACCC CGCAGACCGT CGGTGTGGG CAGCTTTCGG GTGTCTCTG GAGGGGAGC 20580
 TGGGCTGGGC CTGTGACTCC TCAGCCTCTG TTTTCCCCA GGGATGTGCG TGGGGGCCAA GGGCGCCGCC 20650
 GGCCCTCTGC CCTCCGAGGC CGTGCAGTGG CTGTGCCACC AAGCATTCTT GCTCAAGCTG ACTCGACACC 20720
 GTGTCACTA CGTGCCACTC CTGGGTCAC TCAGGACAGG CAAGTGTGGG TGGAGGCCAG TCGGGGCCCC 20790
 ACCTGCCCAG GGGTCACTCT TGAACGCCCT GTGTGGGGCG AGCAGCTCA GATGCTGCTG AAGTGCAGAC 20860
 GCCCCCGGGC CTGACCTTGG GGGCCTGGAG CCACGCTGGC AGCCCTATGT GATTAAACGC TGGTGTCCCC 20930
 AGCCACGGA GCCTGGCAGG GTCCCAACT TCTTGAACCC CTGCTTCCCA TCTCAGGGG GATGGCTCCC 21000
 CACGCTTGGG AGCCTTCTGA CCCCTGACCT GTGTCTCTC ACAGCCTCTT CCCTGGCTGC TGCCCTGAGC 21070
 TCCTGGGGTC CTGAGCAAGT TCTCTCCCG CCCCGCGCT CAGCGCTCAC TGGGCTGCTT GTCTGCTGC 21140
 CCCGGTGGAG GGGTGTCTGT CCCTTCACTG AGGTTECCAC CAGCCAGGGC CACGAGGTGC AGGCCCTGCC 21210

5
 10
 15
 20
 25
 30
 35
 40
 45
 50
 55

```

TGCCCGGCCA CCCACACGTC CTAGGAGGGT TGGAGGATGC CACCTCTGGC CTCTTCTGGA ACGGAGTCTG 21280
ATTTTGGCCC CGCAGCCGAG ACGCAGCTGA GTCCGAAGCT CCCGGGGACG ACCTGACTG CCCTGGAGGC 21350
CGCAGCCAAC CCCGCACTGC CCTCAGACTT CAAGACCATC CTGGACTGAT GGCCACCCGC CCACAGCCAG 21420
GCCGAGAGCA GACACCAGCA GCCCTGTCTC GCCGGGCTCT ACGTCCCAGG GAGGAGGGGG CGGCCACAC 21490
CCAGGCCCGC ACCGCTGGGA GTCTGAGGCC TGAGTGAGTG TTTGGCCGAG GCCTGCATGT CCGGCTGAAG 21560
GCTGAGTGTC CGGCTGAGGC CTGAGCGAGT GTCCAGCCAA GGGCTGAGTG TCCAGCACAC CTGCCGTCTT 21630
CACTTCCCA CAGGCTGGCG CTCGGCTCCA CCCAGGGGCC AGCTTTTCCT CACCAGGAGC CGGCTTCCA 21700
CTCCCCACAT AGGAATAGTC CATCCCCAGA TTGCCCATTTG TTACCCCTC GCCCTGCCCT CCTTTGGCTT 21770
CCACCCCCAC CATCCAGGTG GAGACCCCTGA GAAGGACCCT GGGAGCTCTG GGAATTTGGA GTGACCAAG 21840
GTGTGCCCTG TACACAGGCG AGGACCCTGC ACCTGGATGG GGGTCCCTGT GGGTCAAATT GGGGGGAGGT 21910
GCTGTGGGAG TAAATACTG AATATATGAG TTTTTCAGTT TTTGAAAAAA TCTCATGTTT GAATCCTAAT 21980
GTGCACTGCA TAGACACCAC TGTATGCAAT TACAGAAACC TGTGAGTGAA CGGGGTGGTG GTCAGTGGCG 22050
GCCCATGGCC TGGCTGTGCA TTTACGGAAG TCTATGAGTG AATGGGGTTG TGGTCAGTGC GGGCCCATGG 22120
CCTGGCTGGG CCTGGGAGGT TTCTGATGCT GTGAGGCAGG AGGGGAAGGA GGGTAGGGGA TAGACAGTGG 22190
GAGCCCCCAC CCTGGAAGAC ATAACAGTAA GTCCAGGGCC GAAGGGCAGC AGGGATGCTG GGGGCCAGC 22260
TTGGCGCGCG GGGATGATGG AGGGCCTGGC CAGGGTGGCA GGGATGATGG GGGCCCCAGC TGGGGTGGCA 22330
GGGGTGATGG GGGGGGCTGG TCTGGGTGGC GGGGAAGATG GGGAAAGCCTG GCTGGGCCCC TCCTCCCTCT 22400
GCCTCCACCC TGCAGCCGTG GATCCGGATG TGCTTCCCTG GTGCACATCC TCTGGGCCAT CAGCTTTTAT 22470
GGAGGTGGGG GGCAGGGGCA TGACACCATC CTGTATAAAA TCCAGGATTC CTCTCTCTGA ACGCCCCAAC 22540
TCAGGTTGAA AGTCACATTC CGCCTCTGGC CATCTCTTAA AGAGTAGACC AGGATTCCTG TCTCTGAAGG 22610
GTGGGTAGGG TGGGGCAGTG GAGGGTGTGG ACACAGGAGG CTTCAGGGTG GGGCTGGTGA TGCTCTCTCA 22680
TCCTCTTATC ATCTCCAGT CTCTCTCTCT ATCTCTTAT CATCTCCCAG TCTCATCTGT TCCTCTCTTA 22750
TCTCCAGTCT TCATCTGTCA TCCTCTTACC ATCTCCAGT CTCTCTCTT ATCTCTTAT CTCTAGTCT 22820
CATCCAGACT TACCTCCAG GGGGGGTGCC AGGCTCGCAG TGGAGCTGGA CATACGTCCT TCCTCAGGCA 22890
GAAGGAACCT GAAGGATTTG AGAGAACAGG AGGGGGGGCT CAGAGGGAGC CAGTCTTGGG GTGAAGAAAC 22960
AGCCCTCTCT CAGAAGTTGG CTTGGGCCAC ACGAAACCGA GGGCCCTGCG TGAGTGGGTC CAGAGCCTTC 23030
CAGCAGGTCC CTGCTGGGCG CTTATGGTAT GGGCGGGTCC TACTGAGTGC ACCTTGGACA GGGCTTCTGG 23100
TTTGAGTGCA GCCCGGACGT GCCTGGTGTG GGGGTGGGGC CTTATGGCCA CTGGATATGG CGTCATTAT 23170
TGCTGTGCT TCAGAGAATG TCTGAGTGAC CGAGCCTAAT GTGTATGGTG GGGCCCAAGT CACAGACTGT 23240
GTGCTAAATG CACTCTGGTG CTTGGAGCCC CCGTATAGGA GCTGTGAGGA AGGAGGGGCT CTTGGCAGCC 23310
GGCCTGGGGG CGCCTTTGCC CTGCAAACTG GAAGGGAGCG GCCCGGGGCG CCGTGGGCGG ACGAGCTCAA 23380
GTGAGAGGTT GGACAGAACA GGGCGGGGAC TTCCAGGAG CAGAGGGCGG TGCTCAGGCA CACCTGGGTT 23450
TGAATCACAG ACCAACAGGT CAGGCCATTG TTCAGCTATC CATCTTCTAC AAAGCTCCAG ATTCCTGTTT 23520
CTCCGGGTGT TTTTGTGA AATTTTACT AGGATTACTT ATATTTTGT CTAAGTATT AGACCTTAA 23590
AAAAGGTATT TGCTTTGATA TGGCTTAACT CACTAAGCAC CTACTTTAT TGTCTGTTT TATTTATTAT 23660
TATTTATTATT ATTAGAGATG GTGTCTACTC TGTCACCCAG GTTGTAGTG CAGTGGCACA GTCATGGCTC 23730
GCTGTAGCCG CAACCCCCCA GGCTCAAGTG ATCCTCCGGC CTCAGCTTCC CAGAGTGCTG GGTATACAGG 23800
TGTGAGCCAC TGCCCTTGCC TGGCACTTTT AAAAACCACT ATGTAAGGTC AGGTCCAGTG GCTTCCACAC 23870
CTGTCACTCC AGTAGTTTGG GAAGCCGAGG CAGAAGGATT GTCTAGGCC AGGAGTTTGA GACCAGCATG 23940
GGTAACATAG GGAGACCCCA TCTTACAAA AAATGCAAAA AGTTATCCGG GCGTGGGGTG CAGCATCTGT 24010
AGTCCAGCT GCTCGGGAG CTGAGTGGGA GGATCGCTTG AGCCCCGGAG GTCATGGCTG CAGTGAGCTG 24080
TGATTGTACC ATCGCATCC AGCCTGGGCA ACAGAGTGAG ACCCTGTCTC AAAAAAAAAA AAAAAAAAAA 24150
AAGGAGAAGG AGAAGAGAAG AAGAAGGAAG AAGGAAGAG AAGAAGGAAG AAGAAGGAAG AAGAAGGAAG 24220
AAGGAGGCT GCTAGTGCT AGGTAGACTG TCAATCTCA GAGCAAAATG AAAATAACAA AGTTTTAAAG 24290
GGAAAGAAAA ACCCCAGCTC TTTGGACTTC CTTAGGCCCTG AACTTCATCT CAAGCAGCTT CCTTCCACAG 24360
ACAAGCGTGT ATGGAGCGAG TGAGTTCAAA GCAGAAAGG AGGAGAAGCA GGCAAGGGTG GAGGCTGTGG 24430
GTGACACCAG CCAGGACCCC TGAAAGGGAG TGGTTGTTTT CCTGCCTCAG CCCACGCTC CTGCCGTTCC 24500
TGCACCTGCT GTAACCGTGC ATGTTGGTGC CAGGTGCCCA CCGTGGGAAG ATGCTGTGCA GGGGGCTTGC 24570
CAAACTTTGG TGGGTTTCAG AAGCCCCAGG CACTTGTGGC AGGCACAATT ACAGCCCTC CCAAGAGTG 24640
CCCACGCTCT TCTCTGGAA CCGTGAATG GTCAACCCG AAGGCAGAGG CTGGTGAAGG GTGAGTGG 24710
AATACCGCT GCCAGTCAG CGATCTTAA GTCATCTCG ATTATCTGGT GGGCTGATA TGCCACAG 24780
GGTCCCTAGA AGTGAGAGAG GGAGGCAGG GAGAGTCAGA GAGGGGACGT GAGAAGGACC ACTGGCCACT 24850
GCTGGCTTTG AGATGGAGGA GGGGGTCCCC AGCCAAGGA TGGGGGCAGC CGCTCCATGC TGGAAAAGCA 24920
AGCAATCTC CCCGTTCTG AGGGCACAG GCCCTGCCCA CGCTCGATT TCAGGCCAGT GGGACCTGTT 24990
TCAGCTTTCC GGCCTCCAGA GCTGTAAGAT GATGCGTTTG TGTTCAAGCA CTAAGCTGCA GTGATTCTG 25060
ACAGCAGCAA ATGGAATAGC AGTACAGGGA AATGAATACA GGCAGAGTTC TCAGAGTGAC TCTCAGCCCA 25130
CCCCGTTGG
  
```

60 Beispiel 5

Der Vergleich der oben beschriebenen genomischen hTC-Sequenz mit der Sequenz
 der hTC-cDNA (Fig. 6; entsprechend SEQ ID NO 2) ermöglichte die Aufklärung der
 Exon-Intron-Struktur des hTC-Gens. Die genomische Organisation des hTC-Gens ist
 in Fig. 7 schematisch dargestellt. Die kodierende Region des hTC-Gens setzt sich aus

16 Exons zusammen, die in ihrer Größe zwischen 62 bp und 1354 bp variieren (s. Tabelle 1). Exon 1 enthält das Translationsstartcodon ATG. Das Translationsstopcodon TGA sowie der 3' untranslatierte Bereich liegen auf Exon 16 (Fig. 8). Ein mögliches Polyadenylierungssignal (AATAAA) wurde weder in Exon 16 noch in den 3195 bp der folgenden 3'-flankierenden Region gefunden. Basierend auf der Konsensussequenz

	5'-Exon				Intron				3'-Exon							
	Prä-mRNA	A/C	A	G		G	T	A/G	A	...	N	C	A	G		G
10	Häufigk.(%)	70	60	80		100	100	95	70			80	100	100		60

wurden die Exon-Intron-Übergänge bestimmt und in Tabelle 1 aufgeführt. Mit Ausnahme der 5'-Splice-Stelle zwischen Exon 15 und Intron 15 stimmen alle Exon-Intron-Übergänge mit der publizierten (Shapiro und Senapathy, 1987) Splice-Konsensussequenz überein. Die Größe der Introns liegt zwischen 104 bp und 8616 bp. Da Intron 6 nur zum Teil isoliert wurde, kann die exakte Länge des hTC-Gens nicht bestimmt werden. Basierend auf der von Intron 6 erhaltenen Teilsequenz von ~4660 bp beträgt die minimale Größe des hTERT Gens 37 kb.

Die Introns 1-5 sowie der 5'-Bereich des Introns 6 sind in Contig 1 enthalten:

Intron 1: bp 11493-11596 (SEQ ID NO 4);

Intron 2: bp 12951-21566 (SEQ ID NO 5);

Intron 3: bp 21763-23851 (SEQ ID NO 6);

5 Intron 4: bp 24033-24719 (SEQ ID NO 7);

Intron 5: bp 24900-25393 (SEQ ID NO 8);

5'-Bereich von Intron 6: bp 25550-26414 (SEQ ID NO 9).

Der 3'-Bereich des Introns 6 sowie die Introns 7-15 sind in Contig 2 an folgenden

10 Positionen lokalisiert:

3'-Bereich von Intron 6: bp 1-3782 (SEQ ID NO 10);

Intron 7: bp 3879-4858 (SEQ ID NO 11);

Intron 8: bp 4945-7429 (SEQ ID NO 12);

Intron 9: bp 7544-9527 (SEQ ID NO 13);

15 Intron 10: bp 9600-11470 (SEQ ID NO 14);

Intron 11: bp 11660-15460 (SEQ ID NO 15);

Intron 12: bp 15588-16467 (SEQ ID NO 16);

Intron 13: bp 16530-19715 (SEQ ID NO 17);

Intron 14: 19841-20621 (SEQ ID NO 18);

20 Intron 15: 20760-21295 (SEQ ID NO 19).

Der 3'-nichttranskribierte Bereich befindet sich ebenfalls im Contig 2 an Position
21960-25138 (SEQ ID NO 20).

25 Die genannten Introns haben im einzelnen folgende Sequenzen:

Intron 1 (SEQ ID NO 4)

GTGGGCCCTCCCCGGGGTTCGGCGTCCGGCTGGGGTTGAGGGCGGCCGGGGGGAACCAGCGACATGCGGAGAGCAGCGCAGG
CGACTCAGGGCGCTTCCCCCGCAG

5

Intron 2 (SEQ ID NO 5)

GTGAGGAGGTGGTGGCCGTGAGGGCCAGGCCCCAGAGCTGAATGAGTAGGGGCTCAGAAAAGGGGGCAGGCGAGCC
CTGGTCTCTCTGTCTCCATCGTCACGTGGGCACACGTGGCTTTTCGCTCAGGACGTGAGTGGACACGGTGATCTCTGCC
TCTGCTCTCCCTCTCTGTCCAGTTTGCATAAACTTACAGAGTTACCTTTCACGTTTTGTATGGACACGCGGTTTCCAGGCGC
CGAGGCCAGAGCAGTGAACAGAGGAGGCTGGGCGCGGCAGTGGAGCCGGGTTGCCGGCAATGGGGAGAAGTGTCTGGAAG
CACAGACGCTCTGGCGAGGGTGCCCTGCAGGTTACCTATAATCCTCTTCGCAATTTCAAGGGTGGGAATGAGAGGTGGGA
CGAGAACCCCTCTTCTGGGGGTGGGAGGTAAAGGGTTTTGCAGGTGCACGTGGTCAGCCAATATGCAGGTTTGTGTTA
AGATTTAATTGTGTGTTGACGGCCAGGTGCGGTGGCTCACGCCGTAATCCCAGCACTTTGGGAAGCTGAGGCAGGTGGA
TCACCTGAGGTGAGGAGTTTGAGACCAGCCTGACCAACATGGTGAAACCTATCTGTACTAAAAATACAAAAATTAGCTG
GGCATGGTGGTGTGTGCCGTGAATCCAGCTACTTGGGAGGCTGAGGCAGGAGAATCACTTGAACCCAGGAGGCGGAGGC
TGCAGTGAGCTGAGATTGTGCCATTGTACTCCAGCCTGGGCGACAAGAGTGAAACTCTGTCTTTAAAAAAAAGTGT
CGTTGATTGTGCCAGGACAGGGTAGAGGGAGGGAGATAAGACTGTTCTCCAGCACAGATCCTGGTCCCATCTTTAGGTAT
GAAGAGGGCCACATGGGAGCAGAGGACAGCAGATGGCTCCACCTGCTGAGGAAGGGACAGTGTTTGTGGGTGTTACGGGG
ATGGTGTCTGTGGGCCCTGCCGTGTCCCCACCCTGTTTTCTGGATTGTATGTTGAGGAACCTCCGCTCCAGCCCCCTT
TGGCTCCCAGTGCTCCCAGGCCCTACCGTGGCAGCTAGAAGAAGTCCCGATTTCACCCCTCCCCACAACTCCCAAGAC
ATGTAAGACTTCCGGCCATGCAGACAAGGAGGGTGACCTTCTTGGGGCTCTTTTTTTCTTTTTTCTTTTTATGGTGGC
AAAAGTCATATAACATGAGATTGGCACTCCTAACACCGTTTTCTGTGTACAGTGCAGAATGTAACTCCGGCGGTGTTA
CAGCAGGTTGCTTGAATGTGCGTCTTGGTACTGGAAGTCCCTACCATCGAACGGCAGCTGCCTCACACCTGCTGC
GGCTCAGGTGGACCACGCGGAGTCAGATAAGCGTCATGCAACCCAGTTTTGCTTTTTGTGCTCCAGCTTCTTCTGTTGAG
GAGAGTTTGAGTTCTCTGATCAGGACTCTGCCTGTCTATGTCTTCTGACTTCAGATGAGGTACAATCTGCCCTGG
CTTATGCAGGGAGTGAGGCGTGGTCCCCGGGTGTCCCTGTACGTGCAGGGTGAGTGAGGCGTTGCCCCAGGTGTCCCT
GTCACGTGTAGGGTGAGTGAGGCGCGGCCCCGGGTGTCCCTGTCCCGTGCAGCGTGATTGAGGTGTGGCCCCGGGTGT
CCCTGTACAGTGTAGGGTGAGTGAGGCGCCATCCCAGGTGTCCCTGTACGTGTAGGGTGAGTGAGGCGTGGTCCCCGG
GTGTCCCTGTCCCGTGCAGGGTGAGTGAGGCACTGTCCCAGGTGTCCCTGTACGTGCAGGGTGAGTGAGGCGCGGTCC
CCGGGTGTCCCTCTCAGGTGTAGGGTGAGTGAGGCGCGGCCCCAGGGTGCTCCCTGTACGTGTAGGGTGAGTGAGGCACC
GTCCCTGGGTGTCCCTCCCAGGTATAGGGTGAGTGAGGCACTGTCCCAGGTGTCCCTGTACGTGCAGGGTGAGTGAGG
CGCGGCCCCGGGTGTCCCTCTCAGGTGCAGGGTGAGTGAGGCGGTGTCCCTGGGTGTCCCTGTCTGCTGAGGGTGAGT
GAGGCTCTGTCCCAGGTGTCTTGGCGTTTGCTCACTTGAGCTTGCTCCTGAATGTTTGCTCTTTCTATAGCCACAGCT
GCGCGGGTTGCCATTGCCTGGGTAGATGGTGAGGCGCAGTGCTGGTCCCCAAGCCTATCTTTCTGATGCTCGGCTCT
TCTTGGTCACTCTCCGTTCCATTTTGCTACGGGGACACGGGACTGCAGGCTCTCGCTCCCGCTGCCAGGCACTGCAG
CCACAGCTTCAGGTCCGCTTGCTCTGTTGGGCTGGCTTGTCTCACCAGTGCCCGCCACATGCATGCTGCCAATACTCC
TCTCCAGCTTGTCTCATGCGGAGGTGGACTCTGGGCTGCCTGTGTCTGTGCCACGTGTTGTCTGGAGACATCCAGAA
AGGGTTCTCTGTGCCCTGAAGGAAAGCAAGTCACCCCAGCCCCCTCACTTGTCTGTTTTCTCCCAAGCTGCCCTCTGC
TTGGCCCCCTTGGGTGGGTGGCAACGCTTGTACCTTAATCTGGGCACCTGCCGCTATTGCTTAGGCTGGGCTCTGCCT
CCAGTCCGCCCTCACATGGATTGACGTCCAGCCACAGGTTGGAGTGTCTCTGTCTGTCTCTGCTCTGAGACCCAGCTG

40

GAGGGCCGGTGTCTCCGCCAGCCTTCGTGAGCTTCCCTCTTGGGTCTTAGTTTTGAATTTCACTGATTTACCTCTGACG
TTTCTATCTCTCCATTGTATGCTTTTTCTTGGTTTATCTTTCAATTCCTTTCTAGCTTCTTAGTTTAGTCATGCCCTTTC
CCTCTAAGTGCTGCCTTACCTGCACCTGTGTTTTGATGTGAAGTAATCTCAACATCAGCCACTTTCAAGTGTCTTAAA
ATACTTCAAAGTGTTAATACTTCTTTAAGTATTCTTATTCTGTGATTTTTTTCTTTGTGCACGCTGTGTTTTGACGTGA
5 AATCATTTTGATATCAGTGACTTTTAAGTATTCTTAGCTTATTCTGTGATTCTTTGAGCAGTGAGTTATTGAACACT
GTTTATGTTCAAGATATGTAGAGTATCAAGATACGTAGAGTATTTAAGTTATCATTTTATTATTGATTTCTAACTCAGT
TGTGTAGTGGTCTGTATAATAACCAATTATTGAAGTTTTCGGAGCCTTGCTTTGTGATCTAGTGTGTGCATGGTTTCCAG
AACTGTCCATTGTAAATTTGACATCCTGTCAATAGTGGGCATGCATGTTCACTATATCCAGCTTATTAAGGTCCAGTGCA
AAGCTTCTGTCTCCTTCTAGATGCATGAAATTTCAAGAAGGAGGCCATAGTCCCTCACCTGGGGGATGGGTCTGTTCAAT
10 TCTTCTCGTTTGGTAGCATTTATGTGAGGCATTGTTAGGTGCATGCACGTGGTAGAATTTTTATCTTCTCTGATGAGTGAA
TCTTTTGGAGACTTCTATGTCTCTAGTAATCTAGTAATTTCTTTTTTAAATTGCTCTTAGTACTGCCACACTGGGCTTCT
TTTGATTAGTATTTTCTGCTGTGTCTGTTTTCTGCCCTTAATTTATATATATATATATATTTTTTTTTTTTTTTTGGAGACA
GAGTCTTGGTCTGTGCGCCAGGGTGAGTGCAAGTGGTGTGATCACAGGTGAGTGAACTTTTACCTTCTGGCCTGAGCCGT
CCTCTCACCTCAGCCTCCTGAGTAGCTGGAACGCAGACACGCCCTACACCTGGCTAATTTTTTAAATTTTTTCTGGA
15 GACAGGGTCTTGCTGTGTTGCCAGGCTGGTCTCAAACCTCTTGACTCAAGGGATCCATCTACCTCGGCTTCCCAAAGTG
CTGAATTACAGGCATGAGCCACCATGTCTGGCCTAATTTTCAACACTTTTATATTCTTATAGTGTGGGTATGTCTGTGTA
ACAGCATGTAGGTGAATTTCCAATCCAGTCTGACAGTCGTTGTTTAACTGGATAACCTGATTATTTTCAATTTTTTGTGTC
ACTAGAGACCCGCTGGTGCATCTGATTCTCCACTTGCCCTGTGTCATGTCTCGTTCCCTGTTTCTCACCACTCTTG
GGTTGCCATGTGCGTTTCTGCGGAGTGTGTGATCCTCTCGTTGCCCTCCTGGTCACTGGGCAATTTGCTTTTATTCT
20 CTTTGCTTAGTGTACCCCTGATCTTTTTATTGTGCTGTGTTGCTTTTGTATTGAGACAGTCTCACTCTGTCAACCA
GGCTGGAGTGAATGGCACAATCTCGGCTCACTGCAACCTCTGCCCTCCTCGGTTCAAGCAGTTCTCATTCCTCAACCTCA
TGAGTAGCTGGGATTACAGGCGCCACCACCGCTGGCTAATTTTTGTATTTTAGTAGAGATAGGCTTTCACCATGT
TGGCCAGGCTGGTCTCAAACCTCTGACCTCAAGTATCTGCCCGCCTTGGCCTCCACAGTGCTGGGATTACAGGTGCAA
GCCACCGTGCCCGGCATACCTTGATCTTTTAAATGAAGTCTGAAACATTGCTACCCTTGTCTGAGCAATAAGACCTT
25 AGTGTATTTAGCTCTGGCCACCCCGAGCCTGTGTGCTGTTTTCCCTGCTGACTTAGTTCTATCTCAGGCATCTTGACA
CCCCCACAAGCTAAGCATTATTAATATTGTTTTCCGTGTTGAGTGTCTCTGTAGCTTTGCCCCCGCCCTGCTTTTCTCTC
TTTGTTCCTGCTGTCTTCTGTCTCAGGCCCGCCGTCTGGGGTCCCTTCTTGTCTCTTTCGTGGTCTTCTGTCTTG
TTATGTCTGGTAAACCCAGCTTTACCTGTGCTGGCCTCCATGGCATCTAGCGACGTCCGGGGACCTCTGCTTATGATGC
ACAGATGAAGATGTGGAGACTCAGAGGAGGGCGGTCTCTTGGCCCGTGAAGTGTCTGGAGCACCAGTGCCAGCGTTC
30 CTTAGCCAGTGAGTGACAGCAACGTCCGCTCGGCTGGGTTGAGCCTGGAACCCAGGCATGTGCGGGTCTGTTGGCT
CCGCGGTGTGAGTTTGAAATCGCGCAACCTGCGGTGTGGCGCCAGCTCTGACGGTGTGCTGCGGGGGAGTGTCTG
CTTCTCCCTTCTGCTTGGGAACAGGACAAGGATGAGGCTCCGAGCCGTTGTGCGCCAACAGGAGCATGACGTGAGCC
ATGTGGATAATTTTAAATTTCTAGGCTGGGCGCGGTGGCTCACGCCCTGTAATCCAGCACTTTGGGAGGCCAAGGCGGG
TGGATCACGAGGTGAGGAGTCTGAGACCATCTGGCCACATGATGAAACCCCATCTGTACTAAAAACAAAAAATTAGC
35 TGGGCGTGGTGGCGGGTGCCTGTAATCCAGCTACTCGGGAGGCTGAGGCAGGAGAATTGCTTGAACCTGGGAGTTGGAA
GTTGAGTGAGCCGACATTGCACCACTGCACTCCAGCCTGGCAACACAGCGAGACTCTGTCTCAAAAAAAAAAAAAA
AAAAAAAAAAAAATCTAGTAGCCACATTAAAAAAGTAAAAAAGAAAAGGTGAAATTAATGTAATAATAGATTTTACTGAA
GCCCAGCATGTCCACACCTCATTTTTAGGGTGTATTGTTGGTGGGAGCATCACTCACAGGACATTTGACATTTTTTGGAGC
TTTGTCTGCGGGATCCCGTGTGTAGTCCCGTGTGGCCATCTCGGCTGGACCTGCTGGGCTTCCATGGCCATGGCT
40 GTTGTACCAGATGGTGCAGGTCCGGGATGAGGTGCGCCAGGCCCTCAGTGAGCTGGATGTGAGTGTCCGGATGGTGCAGG
TCTGGGATGAGGTGCGCCAGGCCCTGCTGTGAGCTGGATGTGTGGTGTCTGGATGGTGCAGGTGAGGGTGGGTCTCCAG

GCCCTCGGTGAGCTGGAGGTATGGAGTCCGGATGATGACAGGTCCGGGGTGAGGTCCGACAGGCCCTGCTGTGAGCTGGATG
TGTGGTGTCTGGATGGTGCAGGTACGGGGTGAGGTCTCCAGGCCCTCGGTAAGCTGGAGGTATGGAGTCCGGATGATGCA
GGTCCGGGGTGAGGTCCGACAGGCCCTGCTGTGAGCTGGATGTGTGGTGTCTGGATGGTGCAGGTCTGGGGTGAGGTACCC
AGGCCCTGCGGTGAGCTGGGTGTGCGGTGTCTGGATGGTGCAGGTCTGGAGTGGAGTCCGACAGCGGTGCCAGACCATGC
5 GGTGAGCTGGATATGCGGTGTCCGGATGGTGCAGGTCTGGGGTGAGGTGCCAGGCCCTGCTGTGAGTTGGATGTGGGGT
GTCCGGATGCTGCAGGTCCGGTGTGAGGTACCCAGGCCCTGCTGTGAGCTGGATGTGTGGTGTCTGGATGGTGCAGGTCT
GGGGTGAAGGTCCGACAGGCCCTGCTGTGAGCTGGATGTGTGGTGTCTGGATGGTGCAGGTCTGGAGTGGAGTCCGACAG
GCCCTCGGTGAGCTGGATGTGAGTGTCCAGATGGTGCAGGTCCGGGGTGAGGTCCGACAGGCCCTGCGGTGAGCTGGATG
TGCGGTGTCTGGATGGTGCAGGTCTGGAGTGGAGTCCGACAGGCCCTCGGTGAGCTGGATGTATGGAGTCCGGATGGTGCC
10 GGTCCGGGGTGAGGTCCGACAGGCCCTGCTGTGAGCTGGATGTGCGGTGTCTGGATGGTACAGGTCTGGAGTGGAGTCCGAC
AGACCTGCTGTGAGCTGGATATGCGGTGTCCGGATGGTGCAGGTACGGGGTGAGGTCTCCAGGCCCTCGGTGAGCTGGA
GGTATGGAGTCCGGATGATGCAGGTCCGGGGTGAGGTCCGACAGGCCCTGCTGTGAAGTGGATGTGCGGCGTCTGGATGGT
GCAGGTCTGGGGTGTGGTCCGACAGGCCCTCGGTGAGCTGGAGGTATGGAGTCCGGATGATGCAGGTCCGGGGTGAGGTCCG
CCAGGCCCTGCTGTGAGCTGGATGTGCGGCGTCTGGATGGTGCAGGTCTGGGGTGTGGTCCGACAGGCCCTCGGTGAGCTG
15 GAGGTATGGAGTCCGGATGATGCAGGTCCGGGGTGAGGTGCCAGGCCCTGCTGTGAGCTGGATGTGCTGTATCCGGATG
GTGCAGTCCGGGGTGAGGTCCGACAGGCCCTGCTGTGAGCTGGATGTGCTGTATCCGGATGGTGCAGGTCTGGGGTGAGGT
CACCAGGCCCTGCGGTGAGCTGGTGTGCGGTGTCCGGTGTCTGCAGGTCCGGGGTGAGTCCGACAGGCCCTCGGTGAGC
TGGATGTGCGGTGTCCCGTGTCCGGATGGTGCAGGTCCAGGGTGAGGTCCGAGGCCCTTGGTGGGCTGGATGTGCCGT
GTCCGGATGGTGCAGGTCTGGGGTGAGGTCCGACAGGCCCTTGGTGAGCTGGATGTGCGGTGTCTGCATGGTGCAGGTCTG
20 GGGTGAGGTCCGACAGGCCCTTGGTGGGCTGGATGTGTGGTGTCCGGATGGTGCAGGTCCGGCGTGGAGTCCGACAGGCCCT
GCTGTGAGCTGGATGTGCGGTGTCTGGATGGTGCAGGTCCGGGGTGAGGTAGCCAAGGCCCTTGGTGAGCTGGATGTGGG
GTGTCCGGATGGTGCAGGTCCGGGGTGAGGTCCGACAGGCCCTGCGGTAGCTGGATATGCGGTGTCCGGATGGTGCAGGT
CCGGGGTGAGGTACCCAGGCCCTGCGGTAGCTGGATGTGCGGTGTCTGGATGGTGCAGGTCCGGGGTGAGGTCCGACAGG
CCCTGCTGTGAGCTGGATGTGCTGTATCCGGATGGTGCAGGTCCGGGGTGAGGTCCGACAGGCCCTGCAGTGGAGCTGGATG
25 TGCTGTATCCGGATGGTGCAGGTCTGGCGTGGAGTCCGACAGGCCCTGCGGTAGCTGGATATGCGGTGTCCGGATGGTGCA
GGTCCGGGGTGAGGTACCCAGGCCCTGCGGTAGCTGGATGTGCGGTGTCCGGATGGTGCAGGTCTGGGGTGAGGTCCGAC
AGGCCCTGCTGTGAGCTGGATGTGCTGTATCCGGATGGTGCAGGTCCGGGGTGAGGTCCGACAGGCCCTGCGGTGAGCTGG
ATGTGCTGTATCCGGATGGTGCAGGTCTGGCGTGGAGTCCGACAGGCCCTGCGGTGAGCTGGATGTGAGTGTACGGATGG
TGCAGGTCCGGGGTGAGGTCCGACAGGCCCTGCGGTGGGCTGTATGTGTGTGTCTGGATGGTGCAGGTCCGGGGTGAGTT
30 CGCCAGGCCCTGCGGTGAGCTGGATGTGTGGTGTCTGGATGTGCAGGTCCGGGGTGAGTCCGACAGGCCCTCGGTGAGC
TGGATATGCGGTGTCCCGTGTCCGAATGGTGCAGGTCCAGGGTGAGGTCCGACAGGCCCTTGGTGGGCTGGATGTGCCGT
GTCCGGATGGTGCAGGTCTGGGGTGAGGTCCGACAGGCCCTTGGTGAGCTGGATGTGCGGTGTCCGGATGGTGCAGGTCCG
GGGTGAGGTACCCAGGCCCTCGGTGATCTGGATGTGGCATGTCTCTCGTTTAAG

35 Intron 3 (SEQ ID NO 6)

GTACTGTATCCCAAGCCAGGCCCTGCTTCTCGAAGTCTGGAACACAGCCCGGCTCAGCATGCGCCTGTCTCCACT
TGCTGTGCTTCCCTGGCTGTGACAGCTCTGGGCTGGGAGCCAGGGGCCCGTCACAGGCCCTGGTCCAAGTGGATTCTGTG
CAAGGCTCTGACTGCTGGAGCTCACGTTCTCTTACTTGTAAAATCAGGAGTTTGTGCCAAGTGGTCTCTAGGGTTTGTG
AAGCAGAAGGGATTAAATTAGATGGAACACTACCACTAGCCTCCTTGCCCTTCCCTGGGATGTGGGTCTGATTCTCTC
40 TCTCTTTTTTTTTCTTTTTTGGAGATGGAGTCTCACTCTGTTGCCAGGCTGGAGTGCAGTGGCATAATCTTGGCTCACT

GCAACCTCCACCTCCTGGGTTTAAGCGATTACCAGCCTCAGCCTCCTAAGTAGCTGGGATTACAGGCACCTGCCACCAC
GCCTGGCTAATTTTGTACTTTTAGGAGAGACGGGGTTTACCATTGTTGGCCAGGCTGGTCTCGAACTCATGACCTCAGG
TGATCCACCCACCTTGGCCTCCCAAAGTGCTGGGTTTACAGGCTAAGCCACCGTGCCAGCCCCGATTCTCTTTAATT
CATGCTGTCTGTATGAATCTTCAATCTATTGGATTAGGTATGAGAGGATAAAATCCACCCACTTGGCGACTCACTG
5 CAGGGAGCACCTGTGCAGGGAGCACCTGGGGATAGGAGAGTTCCACCATGAGCTAACTTCTAGGTGGCTGCATTGGAATG
GCTGTGAGATTTTGTCTGCAATGTTCCGGCTGATGAGAGTGTGAGATTGTGACAGATTCAAGCTGGATTGTCATCAGTGAG
GGACGGGAGCGCTGGTCTGGGAGATGCCAGCCTGGCTGAGCCCAGGCCATGGTATTAGCTTCTCCGTGTCCCGCCAGGC
TGACTGTGGAGGGCTTTAGTCAGAAGATCAGGGCTTCCCCAGCTCCCCTGCACTCGAGTCCCTGGGGGGCCTTGTGAC
ACCCATGCCCCAAATCAGGATGTCTGCAGAGGGAGCTGGCAGCAGACCTCGTCAGAGGTAACACAGCCTCTGGGCTGGG
10 GACCCGACGTGGTGTGGGGCCATTTCCTTGCATCTGGGGAGGGTCAGGGCTTTCCTGTGGGAACAAGTTAATACAC
AATGCACCTTACTTAGACTTTACAGTATTTAATGGTGTGCGACCCAAACATGGTCATTTGACCAGTATTTTGGAAAGAAT
TTAATTGGGGTGACCGAAGGAGCAGACAGACGTGGTGGTCCCCAAGATGCTCCTTGTCACTACTGGAGCTGTGTTCTG
CCTGGGGGGCCTTGGAGGCCCTCCTCCCTGGACAGGGTACCGTGCCTTTTCTACTCTGTGGGCTGCGGCCTGCGGTG
AGGGCACCAGCTCCGGAGCACCCGCGGCCCACTGTCCACGGAGTGCCAGGCTGTGAGCCACAGATGCCAGGTCCAGGT
15 GTGGCCGCTCCAGCCCCGTGCCCCATGGGTGGTTTGGGGGAAAAGGCCAAGGGCAGAGGTGTGAGGAGCTGGTGGG
CTCATGAGAGCTGATTCTGCTCCTTGGCTGAGCTGCCCTGAGCAGCCTCTCCCGCCTCTCCATCTGAAGGATGTGGCT
CTTTCTACCTGGGGTCTGCTCTGGGGCCAGCCTTGGGCTACCCAGTGGCTGTACAGAGGGACAGGCATCCTGTGTGG
AGGGGCATGGGTTACGTGGGCCAGATGCAGCCTGGGACCAGGCTCCCTGGTGTGATGGTGGGACAGTCAACCTGGGG
GTTGACCGCCGACTGGGCGTCCCCAGGGTTGACTATAGGACCAGGTGTCCAGGTGCCCTGCAAGTAGAGGGGCTCTCAG
20 AGGCGTCTGGCTGGCATGGGTGGACGTGGCCCCGGGCATGGCCTTCAGCGTGTGCTGCCGTGGGTGCCCTGAGCCCTCAC
TGAGTCGGTGGGGCTTGTGGCTTCCCGTGAAGCTTCCCCCTAGTCTGTGTGCTGGCTGAGCAAGCCTCCTGAGGGGCTCT
CTATTGCAG

Intron 4 (SEQ ID NO 7)

25 GTGGCTGTGCTTTGGTTTAACTTCCTTTTAAACAGAAAGTGCCTTTGAGCCCCACATTTGGTATCAGCTTAGATGAAGGG
CCCGGAGGAGGGGCCACGGGACACAGCCAGGGCCATGGCACGGCGCCAAACCATTTGTGCGCACAGTGAGGTGGCCGAGG
TGCCGGTGCTCCAGAAAAGCAGCGTGGGGGTGTAGGGGGAGCTCCTGGGGCAGGGACAGGCTCTGAGGACCACAAGAAG
CAGCCGGGCCAGGGCCTGGATGCAGCACGGCCCCAGGTCTCGATCCGTGTCTGTGCTGGTGCAGCCTCCGTGCGCT
TCCGCTTACGGGGCCCCGGGGACAGGCCAGACTGCCAGGAGCCACCGGGCTCTGAGGATCCTGGACCTTGGCCACGG
30 CTCTGCACCCACCCCTGTGGCTGCGGTGGCTGCGGTGACCCCGTCATCTGAGGAGAGTGTGGGTGAGGTGGACAGAG
GTGTGGCATGAGGATCCCGTGTGCAACACACATGCGGCCAGGAACCCGTTTCAAACAGGGTCTGAGGAAGCTGGGAGGGG
TTCTAGGTCCCGGTCTGGGTGGCTGGGGACACTGGGGAGGGGCTGCTTCTCCCTGGGTCCCTATGGTGGGTGGGCAC
TTGGCCGGATCCACTTTCCTGACTGTCTCCATGCTGTCCCCGCCAG

35 Intron 5 (SEQ ID NO 8)

GTGGGTGCGGGGACCCCGTGAGCAGCCCTGCTGGACCTTGGGAGTGGCTGCCTGATTGGCACCTCATGTTGGGTGGAG
GAGGTACTCCTGGGTGGGCGCAGGGAGTGCAAGTGACCTGTCACTGTTGAGGACACACCTGGCACCTAGGGTGGAGGC
CTTCAGCCTTTCTGCAGCAGATGGGGCCGACTGTGCACCTGACTGCCCGGGCTCCTATCCCAAGGAGGGTCCACTG
GATTCCAGTTTCCGTGAGAGAAGGAACCGCAACGGCTCAGCCACCAGGCCCGGTGCCTTGACCCACAGTCTGAGCCAG
40 GGGTCTCCTGTCTGAGGCTCAGAGAGGGGACACAGCCCGCCTGCCCTTGGGGTCTGGAGTGGTGGGGTGCAGAGAGAG

5'-Bereich Intron 6 (SEQ ID NO 9)

15

TGTGGGATTGGTTTTTCATGTGTGGGATAGGTGGGGATCTGTGGGATTGGTTTTTATAGTGGGGTAACACAGAGTTCAGG
 GCGAGCTTTCTTCTGTAGTGGGTCTGCAGGTGCTCCAACAGCTTTATTGAGGAGACCATATCTTCTTTGAACTATGGT
 CGGGTTTATAGTAAGTCAGGGGTGTGGAGGCCTCCCTGGGCTCCCTGTTCTGTCTTCTTCCACTCTGGGGTCGTGTGGT
 CCTGCTGTGGTGTGTGGCCGGTGGGCAGGGCTTCCAGGCCTCCTTGTTGTTTCATTGGCCTGGATGTGGCCCTGGCTACGCT
 CCGTCTTGGGAATCCCTGCGAGTTGGAGGCTTCTTCTTCTTCTTTTTCTTCTTCTTTTTTTTTTTTGATAACAGA
 GTCTCGCTCTTTTTTGCCAGGCTGGAGTGGTTGGCGTGATCTTGGCTCACTGCAACCTGTGCTTCTCTGAGTTCAAGCA
 ATTCTCTTGCTCAGCCTCCCAAGTAGCTGGAATTATAGGCGCCACCACCATGTGACTAATTTTTGTAAATTTAGTAG
 AGACGAGGTTTCTCCATGTTGGCCAGGCTGGTCTCGAACTCCTGACCTCAGGTGATCCTCCACCTCGGCCCTCCCAAAGT
 GCTGGGATGACAGGTGTGAACCGCGCGCCGGCCGAGACTCGTCTCTGCAGCTTCCGTGAGATCTGCAGCGATAGCTG
 CCTGCAGCCTTGGTGTGACAACTCCGTTTTCTTCTCCAGGTCTCGCTAGGGGTCTTCCATTTCTGACTCTCTTCA
 CAGAAGAGTTTACGTGTGCTGATTCCCGGCTGTTTCTCGCTAATTGGTGTCTGCTGTTTATCGATGGCCTCCTTCCA
 TTTCTTCTTAGGCTTGTGTTATGTGTTTTTCCGGCTCCTTGAAGGAAAAGTTTCGATTATGGATGTTTGAACCTTCTT
 TCTAACAAGCATCTGAAGTTGCCGTTTTCCCTCTAAGCAGGGATCCCGAGGCCCTGGCTGTGGAGTGGCACC GGTC
 GGGGCTGTTAGGAACCCGGCGCACAGCGGGAGGCTAGGTGGGGTGTGGGGAGCCAGCGTTCCCGCCTGAGCCCCGCCCC
 TCTCAGATCAGCAGTGGCATGCGGTGCTCAGAGGCGCACACCCCTACTGAGAACTGTGCGTGAGAGGGGTCTAGATCT
 GTGCTCCTTATGGGAATCTAATGCCTGATGATCTGAGGTGGAACCGTTTGTCTCCAAAACCATCCCCTTCCCCACTGCT
 TCTGTGGAAAAATCGTCTTCCACGAAACCAAGTCCCTGGTACCACAAATGGTTGGGGACCTGTGCTAAAGACCTGCTTCA
 GCAGCCTCTCGTCAGTGTGATATATGGCTTTTCTGTGTTGAGTCCAGAATAATTACGGATTCTGTGATGCTTTCCG
 CGACCTCAGACCCATGGGCTATTTGTGGGCGTGTGCCTGCTCCTGGGTGGGAAGGGTGCAAGGCCCATGTACCTTCT
 GTTACTGCCTTCCAGGTTGGTTCTCAGGTTGAATCGTACTCGATGTGTTTTAGCCACAGGCCCTGCGCCAGCTCCT
 GGGGCTGGGGAACATGCTGAAGCACAGAGTACCCTGCGCGTCTTTTGATGCCTACAAGCTCGAGGCCCTCTGTGTCC
 TGTAGTGTGTGTCACGTGCCTGCTCACATCCTGTCTTGGGGACGCAGGGGCTTAGCAGGTCCCGTAGTAAATGACAAG
 GTCCTGGGGGAGTCTGCAGAATAGGAGGTGGGGGTGCGGCTCTCTCTCCCGCGTCTTCAGACTCTTCTCCTGCGCTGTG

44

GTGGCTGCACCTGCATCCCTGCAATCCCTCCAGCACTGGGCTGGAGAGGCCCGGAGCTCGAGTGCCACTTGTGCCACGT
GACTGTGGATGGCAGTCGGTCACGGGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTTGGTCACAGGGGTCTGATGTGTG
GTGACTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGAAGTGTGG
ATGGCGGTCGTGGGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCGTG
5 GGGTCTGATGTGTGGTGAAGTGTGGATGGCAGTCGTGGGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCGTGGGGTCTGATG
TGGTGAAGTGTGGATGGCAGTCGTGGGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGAAGT
GTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGAAGTGTGGATGG
CGGTCGTGGGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGAAGTGTGGATGGTGAAGTCGTC
CAGGGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGAAGTGTGGATGGTGAAGTCGTCACAG
10 GGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCGTGGGGTCTG
TCTGATGTGTGGTGAAGTGTGGATGGCGATCGGTCACAGGGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCGTGGGGTCT
GATGTGTGGTGAAGTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGT
GACTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGAAGTGTGGAT
GGCGGTCGTGGGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGAAGTGTGGATGGCAG
15 TCGTGGGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCGTGGGG
TCTGATGTGTGGTGAAGTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCGTGGGGTCTGATGT
GGTGAAGTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGAAGTGTGGATGGTGAAGTCGTCACAGGGGTCTGATGTGTGGT
GACTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGAAGTGTGGAT
GGCGGTCGTGGGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGAAGTGTGGATGGCAGTCG
20 GTCACAGGGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCGTGG
GGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCGTGGGGTCTGAT
GTGGTGAAGTGTGGATGGTGAAGTCGTCACAGGGGTCTGATGTGTGGTGAAGTGTGGATGGAGTCCCAGGTGTGTCTGTAGCT
ACTTTGGCTCCTCGGCCCCCGGCCCCGTTTCCCAAAACAGAAGCTTCCAGGCGCTCTCTGGGCTTCATCCCGCCATCG
GGCTTGGCCGAGGTCCACAGTCTGATCGGAAGAAACAAGTCCCCAGCTCTGGCCGGGAGGCCACATTTGTGGGCTC
25 ATGCCCTCTCCTCTGCCGCGAG

Intron 7 (SEQ ID NO 11)

GTCTGGGCACTGCCCTGCAGGGTTGGGCACGGACTCCAGCAGTGGGTCCTCCCCTGGGCAATCACTGGGCTCATGACCG
GACAGACTGTGGCCCTGGGGGGCAGTGGGGGAATGAGCTGTGATGGGGGCATGATGAGCTGTGTGCCCTTGGCGAAATC
30 TGAGCTGGGCCATGCCAGGCTGCGACAGCTGCTGCATTAGGCACCTGCTCACGTTTGAAGTGCAGGCTCTCTCCAGTT
CCGCAAGTGCCTTTGTTTATGATTGCTAAATGCTCTCTGCCAGTTTGTATCTTGAAGCCAAAGGAAAGGTGTCCCCCT
CCTTTAGGAGGGCAGGCCATGTTTGAAGCGTGTCTGCCCCAGCTGGCCCTCAGTGTGGGCTGAGGCCAAAGGAAACG
TGTCCCCCTTCTTAGGAGGACGGGCGTGTGAGCCAGCCCCGCTGAGCGGGCTCTCAGTGTGGGCTGTGTCCACGT
GGCCCTGTGGCCCTTTCAGATGTGGTCTGTCCACGTGGCCCTGTGGCTCTTTCAGATGCCTGTAGCACTTGTCTGGC
35 TCTAGGGGACAGTCTGTCTCCACCATGAGGCTCAGAGACCTCTGGGCGAATTCCTTGGCTCCCAGGGTGGGGGTGGAG
GTGGCCTGGGCTGTGGGACCCAGACCTGTGCCCGGAGCTGGGCAGCAACTCCTGGATCACATATGCCATCCGGGCCA
CGGTGGGCTGTGTGGGTTGAGCCAGCTGGACCCACAGGTGGCCAGAGGAGACGTTCTGTGTACACACTCTGCCTAA
GCCCATGTGTGTCTGACAGAGACTCGGCCCGGCCAGCCACAGATGGCCCTGCATTCCAGCCAGCCCGCACTTCATCACA
AAGTGTGACCCAAAGGAGCGAGGGTCTTGGCCACGTGGTCTGCTGTCTCAGACCCACCGGCTCACTCCCATGTG
40 TCTCCCGTCTGCTTTCGCGAG

Intron 8 (SEQ ID NO 12)

GTGAGTCAGGTGGCCAGGTGCCATTGCCCTGCGGTGGCTGGGCGGGCTGGCAGGGCTTCTGCTCACCTCTCTCCTGCCCC
CTTCCCCACTGNCCTTCTGCCCCGGGGCCACCAGAGTCTCCTTTTCTGGCCCCCGCCCCCTCCGGCTCCTGGGCTGCAGGC
5 TCCCCAGGCCCCCGAAACATGGCTCGGCTTGGCGAGCCGGAGCGGAGCAGGTGCCACACGAGGCCTGGAAATGGCAAGC
GGGGTGTGGAGTTGCTCCTGCGTGGAGGACGAGGGGCGGGGGTGTGTCTGGGTGAGGTGTGCGCCGAGCGTTTGAGCCT
GCAGCTTGTGAGCTCCAAGTTACTACTGACGCTGGACACCCGGCTCTCACACGCTTGTATCTCTCTCTCCCGATACAAAA
GGATTTTATCCGATTCTCATTCTGTCCCTGTGCTGTGACCCCGGAGGGCGCGGGCTCTTCTCTGTGACTAGATTT
CCCATCTGGAAAGTGCGGGGTTGACCGTGTAGTTTGCTCCTCTCGGGGGGCTGTGGTGCCATGGGGCAGGCGGCCTGG
10 GAGAGCTGCCGTACACAGCCACTGGGTGAGCCACACTCACGGTGGTAGAGCCACAGTGCCTGGTGCCACATCACGTCTT
CTGGATTTTAAGTAAAACACACACCTCCCGGCAGGCATCTGCCTGCGACCCTGTGTGTGCTGGGGAGAGTGGTAGCAC
GGAGGAAATTCGTGCACACTCAAGGTTCATCAGCAAGGTTCATCCGCAGTCAGGTGGAACGTGGAGGCCTCTCTCTGGGATC
GTCTCCAGCGGATAAAGGACTGTGCACAGCTTCGGAAGCTTTTATTTAAAAATATAACTATTAAATTATTGCATTATAAGT
AATCACTAATGGTATCAGCAATTATAATATTATTAAAGTATAATTAGAAATATTAAGTAGTACACACGTTCTGGAAAAA
15 CACAAATTCACATGGCAGCAGAGTGAATTTTGGCCGAGGGACACGTGTGCACATGTGTGTAAGCGGCCCCCAGGCCAC
AGAATTCGCTGACAAAGTCACCTCCCCAGAGAAGCCACCAGGGCCTCCTTCGTGGTGTGAAATTTTATTAAGATGGATC
AAGTCACGTACCGTCCACGTGTGGCAGGGCTTTGGGGAATGTGAGGTGATGACTGCGTCTCATGCCCTGACAGACAGGA
GGTGACTGTGTCTGTCTGTCCCTAGGACACGGACAGGCCCGAAGCTCTAGTCCCATCGTGGTCCAGTTTGGCCTCTGA
ATAAAAACGTCTTCAAAACCTGTTGCCCAAACTAAGAACAGAGAGAGTTCCCATCCCATGTGCTCACAGGGCGGTA
20 TCTGCTTGGTTGACTCGCTGGGCTGGCCGACTCCTAGAGTTGGTGGTGTGCTTCTGTGCAAAAGTGCAGTCTCTT
GCCCCATCACTGTGATATCTGCACCAGCAAGGAAAGCCTCTTTCTTTCTTTCTTTCTTTTCTTTTGGAGACGGAAACGTCA
CTGTGTCTGCTGGGCTTGAGTGCAGTGGCGCATCTCAACTCACTGCAACCTCCGCCTCCCGGGTTCAGCATTTCTC
CTGCCCTCAGCCTCCCGAGCAGCTGAGATTACAGGCACCCACCCCTGCGCTGGCTAATTTTGTATTTTGTAGTAGAG
GGGTTTTTGCCATGTTGGCCAGGCTGGTCTCGAACTCCTGACCTCAGGTGATCCACCCACCTCGGCCTCCCAAAGTGTCTG
25 GGATTACAGGTGTGAGCCATCACGCCCAGCCGAAAGCCTCTTTTAAAGTGACCACTATAGCGCTTCCCGAAATAAC
AGGTCTTGTTTTTCAGTAGGCTGCAAGCGTCTCTTAGCAACAGGAGTGGCGTCTGTGGGCTCTGGGGATGGCTGAGGG
TCGCGTGGCAGCCATGCCTTCTGTGTGACCTTTAGGTTCCACGGGGCTATTCTGCTCTCACTGTTTGTCTGAAAACGCA
CCCTTGGCATCCTTGTGTTGGAGAGTTTCTGCTTCTCGTTGGTCATGCTGAAACTAGGGGCAAGGTTGTATCCGTTGGCGC
GCAGCGGTACATGTAGGTCATGAGTCTTTACCGTGGACAAATCCTTGAAAAAAGAGGAGTCCGGTTAAGCAT
30 TCATTCCGGGTCAAGTGTCTGGTTCTGTGAATAAACTCTAAGATTAAAGAAACCTTAATGAAAGAAACCTTGATGATTC
AGAGCAAGGATGTGGTCACACCTGTGGCTGGATCTGTTTACGCCGCCAGTGTCATGGTGAGAGTGGGGAGCAGGGATTG
TTGTGTCAGAGGTCTCATCTGGTATGTTTCTGAGGTGTTTGGCGGTGAATGGTAGACGTGCTGTTGTGTATGAGGT
TCTGTGTCTGTGTGGCTCGGTTTGTAGGTACGCATGTCCAGCACATGCCCTGCCCGTCTCTCACCTGTGTCTTCCCGC
CCAG

35

Intron 9 (SEQ ID NO 13)

GTGAGGCCTCCTCTTCCCCAGGGGGCTTGGGTGGGGTTGATTGCTTTTGATGCATTCAAGTGTAAATATTCCTGGTGC
TCTGGAGACCATGACTGCTGTCTTGGAGAACAGACAAGGTTGCAGCCCCCTTCTGGTATGAAGCCGACGGGAGGGG
TTGCACAGCCTGAGGACTGCGGGCTCCACGCAGGCTCTGTCCAGCGGCCATGTCCAGAGGCCTCAGGGCTCAGCAGGCGG
40 GAGGGCCGCTGCCCTGCATGATGAGCATGTGAATTCAACACCGAGGAAGCACACAGCTTCTGTACGTACCCAGGTTT

CGTTAGGGTCCTTGGGGAGATGGGGCTGGTGCAGCCTGAGGCCCCACATCTCCAGCAGGCCCTCGACAGGTGGCCTGGA
CTGGGGCGCTCTTCAGCCCATGCCCATCCCCTGTCATGGGGCTACACCCAGGACGCACACCTAAATATCGTGCC
AACCTAATGTGGTTCAACTCAGCTGGCTTTTATTGACAGCAGTTACTTTTTTTTTTTAATACTTTAAGTTCTAGGGTAC
ATGTGCACGACGTGCAGGTTAGTTACATATGTATACATGTGCCATGTTGGTGTGCTGCACCCATTAACTCATCATTTACA
5 TTAGGTATATCTCCTAATGCTATCCCTCCCCCTCCCCCATCCCATGACAGGCCCTGGTGTGTGATGTTCCCAACCTG
TGTCGAAGTGTTCTCATTGTTCAAGTCCCACCTGTGAGTGAGAACATGTGGTGTGTTGGTCTTCTTCTTCAATAGTTT
GCTCAGAGTGATGGTTCCAGCTTCGTCCATGTCCCTACAAAGGACATGAACTCATCCTTTTTTATGACTGCATAGTATT
CCGTGGTGTATATGTGCCACATTTTCTTAATCCAGTCTATCATCGATGGACATTTGGGTTGGTTGCAAGTCTTTGCTACT
GTGAATAGTGCCGCAATAACATACGTGTGCATGTGTCTTTATAGCAGCATGATTATAATCCTTTGGGTATATACCCAG
10 TAATGGGATGGCTGGGTCAAATGGTATTTCTAGTTCTAGATCCTTGAGGAATCACCACACTGTCTTCACAATGGTTGAA
CTAGTTTACACTCCCAACAGTGTAAAAGTGTCTGGTGTGAGAGGATGTGGACAGCAGTTATTTTTTATGAAAA
TAGTATCACTGAACAAGCAGACAGTTAGTGAAGGATGCGTCAGGAAGCCTGCAGGCCACACAGCCATTTCTCTGAAGAC
TCCGGGTTTTTCTGTGCATCTTTGAAACTCTAGCTCCAATTATAGCATGTACAGTGGATCAAGGTTCTTCTTCATTAA
GGTTCAAGTTCTAGATTGAAATAAGTTTATGTAAACAGAAACAAAAATTTCTGTACACAACTGTCTGGGATTGGA
15 GGAAAGTGTCCTCGAGCTGGCGGCACACTGGTCAGCCCTCTGGGACAGGATACCTCTGGCCCATGGTCATGGGCGCTGG
GCTTGGGCTGAGGGTCACACAGTGCACCATGCCAGCTTCTGTGGATAGGATCTGGGTCTCGGATCATGCTGAGGACC
ACAGCTGCCATGCTGGTAAAGGGCACCAGTGGCTCAGAGGGGGCAGGTTCCAGCCCCAGCTTTCTTACCGTCTTCAG
TTATTTTTCCCTAAGAGTCTGAGAAGTGGGGCCGCGCTGATGGCCTTGGTTCGTCTTCAGCTGGCACAGAATTGCACAA
GCTGATGGTAAACACTGAGTACTTATAATGAATGAGGAATTGCTGTAGCAGTTAACTGTAGAGAGCTCGTCTGTGGAAA
20 GAAATTTAAGTTTTTCATTTAACCGCTTTGGAGAAATGTTACTTTATTTATGGCTGTGTAATGTTTGACATTCACTCCC
TCGTAGACAGATACTACGTAAAAAGTGAAGTTAACCTTGCTGTGATTTTCCCTTATTTTAG

Intron 10 (SEQ ID NO 14)

GTGAGGCCCGTGCCGTGTCTGTGGGACCTCCACAGCCTGTGGGCTTGCAGTTGAGCCCCCGTGTCTGCCCCCTGG
25 CACCGCAGCGTTGTCTCTGCCAAGTCTCTCTCTGCGGTGCTGGATCCGCAAGAGCAGAGGCGCTTGGCCGTGCACC
CAGGCTGGGGGCGCAGGGGCACCTTCGGGAGGGAGTGGGTACCGTGCAGGCCCTGGTCTCTGAGAGACGCACCCAGGTT
ACACACGTGGTGAGTGCAGGCGGTGACCTGGCTCCTGTGCTCTTTGGAAAGTCAAGAGTGGCGGCTCTGGGGCCCCAG
TGAGACCCCCAGGAGCTGTGCACAGGGCTGCAGGGCCGAGGCGGCAGCCTCCTCCCCAGGGTGACCTGAGCCTGCGGA
GAGCAGGAGCTGCTGAGTGAGCTGGCCACAGCGTTGCTGCGGTACGTTCTGCGTGGGGTTGTTGGGATCGGTGGG
30 AGAATTTGGATTGCTGAGTGTCTGTCTTGAACACGGAGATGGCTAGGAGTGGGTTTCAAGATTGATTTTGTGAAT
CAAATAAAATCAGGCACAGGGGACCTGGCCTCAGCACAGGGGATTGTCCAATGTGGTCCCCCTCAAGGGCGCCCCACAG
AGCCGGTGGGCTTGTTTTAAAGTGCATTTGACGAGGGACGAGAAACCTTGAAAGCTGTAAAGGGAACCTCAGAAAAATG
TGGCCGACAGGGGTGGTTTCAGGTGCTTTGCTGGGCTGTGTTTGTGAAAACCCATTGGACCCGCTCCAAGTCCACCC
TCCAGGTCCACCTCCAGGGCCGCTGGGCTGGGGGTATGCTGCGGTTCTTGTGCGCAGCCCCGAGCACAGCAGGC
35 TGTGCACATTTAAATCCACTAAGATTCACTCGGGGGAGCCAGGTCCCAAGCAACTGAGGGCTCAGGAGTCTGAGGCT
GCTGAGGGGACAGAGCAGACGGGAACGTGCTTCTGTGTGGCAAGTTCTGAGGGTGTGGCCAGGAGGTGGCTCAGA
GTGTATGTTGGGTCCTCCCGGGGGCAGAACTCTGTCTGTATGAGTGGCAGCCATGTAAAGGAAGGGGTGGCCACAG
GGAGCTGGGAATGCACAGGGGAGCTGCGCAGCTGGCCGAGGTCCAGGGCCAGGCCACAGGAAGGGCAGGGGGACGCCC
GGGGCCACAGCAGAGGCCGAGGAAGGAAGGGATGCCAGGCCAGAGCAGAGGCTACCGGGCACAGGGGGCTCCCTG
40 AGCTGGGTGAGCAGGCTCATGACTCGGCAGGGAACCTCCTTGACGTGAAGCTGACGACTGGTGTGCCAGCTCACAG

CCCAGCCAGGTCCCGCGCCTGAGCAGGAACTCAGAACCCTCCCTTTGTCTAAAGCACAGCAGATGCCTTCAGGGCATCT
AGGAGAAAAACAGGCAAAGTCGTTGAGAAAACGTCCTAAAAGAAGGTGGGATGGTGGCAATTTCTGTCCAGATTTTAGTCT
GCCCCGGACCACAGATGAGTCTATAACGGGATTGTGGTGTGGCATGGGGACACATGAGATGGACCATCACAGAGGCCAC
TGGGGCTGCACCTCCCATCTGAGTCCTGGCTGTCCCGGGTCCAGGCCAGGTTCTTGCACTGCTCACCTACCTGTCTGTCCC
5 GGGAGACAGGGAAAGCACCCCGAAGTCTGGAGCAGGGCTGGGTCCAGGCTCCTCAGAGCTCCTGCCAGGCCAGCACCTT
GCTCCAAATCACCACCTTCTCTGGGGTTTTCCAAAGCATTAAACAAGGGTGTGAGTTACCTCCTGGGTGACGGCCCCGCA
TCCTGGGGCTGACATTGCCCTCTGCCTTAG

Intron 11 (SEQ ID NO 15)

10 GTGAGCGCACCTGGCCGGAAGTGGAGCCTGTGCCCGGTGGGGCAGGTGTGCTGCAGGGCCGTTGCGTCCACCTCTGCT
TCCGTGTGGGGCAGGCGACTGCCAATCCCAAAGGTGAGGGCCACAGGGTGCCCTCGTCCCATCTGGGGCTGAGCAGA
AATGCATCTTTCTGTGGAGTGAGGGTGCTCACAACGGGAGCAGTTTTCTGTGCTATTTTGGTAAAAGGAAATGGTGAC
CAGACCTGGGTGCACTGAGGTGTCTTCAGAAAGCAGTCTGGATCCGAACCCAAGACGCCGGGCCCTGTGTGGCGTGAGT
CTCTCAAACCCGAACACAGGGGCCCTGTGGGCATGAGTCCCTCTGAACCCGAGACCTGGGGCCCTGTGTGGCGTGAGT
15 CTCTCCGAACCCAGAGACTTCAGGGCCCTTTTGGGCGTGAGTCTCTCCGCTGTGAGCCCCACACTCCAAGGCTCATCCAC
AGTCTACAGGATGCCATGAGTTCATGATCACGTGTGACCCATCAGGGGACAGGGCCATGGTGTGGGGGGGTCTCTACAA
AATTCTGGGGTCTGTGTTCCCGAGAGCCGAGAGCTCAAGGCCCGTCTCAGGCTCAGACACAATGAATTGAAGATGGA
CACAGATGCAGAAATCTGTGCTGTTCTTTTATGAATAAAAAGTATCAACATTCCAGGCAGGGCAAGGTGGCTCACACCT
ATAATCCAGCACTTTGGGAGGCCGAGGTGGGTGGATCACTTGAGGCCAGGAGTTTGAGGCCAACCTAACCAACATAGTG
20 AAATTCCATTTCTACTTAAAAAATACAAAAATTAGCCTGGCCTGGTGACACGCCTGTAGTCCCGCTATGCGGGAGGC
TGAGGCAGGAGAATCATTTGAACCCAGGAGGCAGAGGTGTCAGTGAGCCGAGATCACACCACTGCACTCCAGCCTGGGCA
ACAGAGTGAGACTTCATCTTAAAAAAAAAAAAAGTATCAGCATTCCAAAACCATAGTGACAGAGTGTTTTTTTATTC
TGTCTTCGATAATATTACTGGTGCTGTGCTAGAGGCCGGAACGGGGTGCCCTCCTCTGAAAGGCACACCTTCATGG
GAAGAGAAATAAGTGGTGAATGGTTGTTAAACCAGAGGTTTAACTGGGGTCTGTGCTGTTCTGAGTTAACAGTCCAGATC
25 TGGACTTTGCCTCTTCCAGAATGCTCCCTGGGGTTTGCTTCATGGGGAGCAGCAGGTGTGGACACCTCGTGATGGGG
GAGCAGCAGGTGCAGACGCCCTCATGATGGGGAGTGGCAGGTGCAGACACCTTGTGATGGTGCCAGCATGTCCCTG
TTGCAGCTCCCTCCCCACAAGGATGCCGGTCTCCTGTGCTCCCCACAGTCCCTGCTTCCCTCTCACAGCCTTACCTGGTC
CTGGCCTCCACTGGCTTTGTCTGCATGATTTCCACATTTCTGGGCTCCAGCACCTCTTCGCTCTCCAGGCACCTCT
GCAGTGCTGGCCATACCAGTCAGCTGTGAAGTGTCCACTGCTTATTTGCTCCCCATGAAATGTATTTTTTAGGACAGGC
30 ACCCTTGGTTCCAGCCTCTGGCACAGCATCAGTGAATGTTATTGAAGGACAAAGGACAGACAAACAAATCAGGAAAATGG
GTTCTCTCTAAACACATTGCAAAGCCACAGAGGCTAGTGACAGGATGGGTGGGCATCAGGTATCAGATGTGGGTCCAATG
CCAGAATATTCTGTGCTCCCAAAGGCCACTTGGTCAGAGTGTGTGCTTGACAGAGGTGGCTCTAAAGCTCAGCAGTGGAG
GCAGTGGTTCGCCATACTCAGGGTGAAGTACATCCTCTGTGCTGTAAGTATACAGCAGAGGCTTGAAGGGCATCTGGGA
GAAGAAAACAGGCAAATGATTAAGAAAAGTGAAGAAAGGAAAGTGGTAAGATGGGAATTTCTGTCCAGATTTTAGTC
35 TCCCAAACACAGCTCAGATGGTAGAATGTGGTCAGAACTGATGGACAGAACATAGAACAAAACGGAAGCCCTATCTCT
CAGAAACGTGTGTTAATGTGGTATGTGGCAGAGCTGATGAAAAGAGAGTGTGTGTGAATTTTTTTCTGAGAAAACT
GACTGGAAGCAAATAAGTTGTGCTTTACAGCATATACCAGAGCAGATTCTAGGTAGAAGAGGAGACATGCAACAAC
ACCAGCAACAGAAATAAAACAAAAGACTCAAAGGGAAGGGAGGTGAACGTTCCCTGGTTTGGTGTGGGGAAGGACACAC
AGGGAGGCGGATGAAACAGTGAGGCAACGGGCATTGCTTTCAGTGCAGAGAACTCAGCTTGCCTGAGCCACAGTGAAA
40 ATGGCCATTCCCTGGAGCGTTTGTGCACGTGATTTATTTAAGGCGCCCTGTGAGGTCTGCACATTCATCCTCTCACTTT

5 GTTCTCCTAACCACTGAGAGGTAGAGGAGAAAGGCTCCAGGGGAGCAGCCGCCCTTGGTCACCCAGCTGGCAAAGGGC
ATGCATGATTGCAGCCTGGCCTCCTGCTCCGGGGCCCTTGCTCTGCCCGAGGACCCACACAAGTCAGACCCATAGGCTC
AGGGTGAGCCGGAGCCCAAGGTCGTGTTGGGGATGGCTGTGAAAGAAGAAATGGACGTCTGATGCACACTTGGGAAGGTC
CTACCAGCAGCGTCAAAGAAATGCATGTGAAACTGACAGCGAGACCCATCCCTCAAAGAAACGCACGTGAAACTGATGGC
10 GAGACCTGTCCCCATCCCTCATGCTGGCTCCTTTTCTGGGCTTGCCAAGAGCCAGCATCAGGTTGAGGCAAGCTGGAAAG
ACTTTTCTGGAAAGCAGCTTGTGTCATGGAAGTCCTCACAATGTCTGTCTTCCAGTAATCCACTTCTGAAGTGA
CCAGACATTATCACGGGTCTTATTTACCATTTCAGTGTTCAGGCAGGGGGACTTGCCACAGCAAGTCACGAACCTGCC
CAAATACAGGGCTAAGGAGATATTATGCATCACAAACTTGCTCTGCCATTAAACATTTTCAAAGAATTTTGAAGAAT
GTTTAATGGCACAAACGTTTATTTCAATGTAGCAGTGTCAAAGCTGGATGTAAAAGAACACACCCAGGAGCCTGCCG
15 TGAATGTCATGTGTTCATCTTTGGACATGGACATACATGGGCAGTGAGTGGTGGTGAGGCCCTGGAGGACATCGGTGG
GATGCCCTCCATCCTGCCCTCTGGAGACACCATGTGTGCCAGTGCACCTCACTGGAGCCCTGTTTAGCTGGTGCCACCTG
GCTCTTCCATCCCTGAGATTCAAACACAGTGAGATTCCCCACGCCCACTCAGTGTCTCCACAAAAAACCCTGAGTCAC
ACCTGTGTTCACTCGAGGGACGCCCGGGAGCCAGGGCTCCACAGTTTATTATGTGTTTTTGGCTGAGTTATGTGCAGATC
TCATCAGGGCAGATGATGAGTGACAAACACGGCCGTGCGAGGTTTGGATACACTCAACATCACTAGCCAGGTCTCGGTG
20 GAGTTTGGTCATGCAGAGTCTGGATGGCATGTAGCATTGGAGTCCATGGAGTGAGCACCCAGCCCTCGGGCTGCAGC
GCATGCCCCAGGCAGGACAAGGAAGCGGGAGGAAGGCAGGAGGCTCTTTGGAGCAAGCTTTCAGGAGGGGGCTGGGTGT
GGGGCAGGCACCTGTGTCTGACATTCCCCCTGTGTCTCAG

Intron 12 (SEQ ID NO 16)

20 GTGAGCAGGCTGATGGTCAGCACAGAGTTCAGAGTTCAGGAGGTGTGTGCGCAAGTATGTGTGTGTGTGTGCGCGCT
GCCTGCAAGGCTGATGGTGACTGGCTGCACGTAAGAGTGACATGTACGCATATACAGTGAGCACATACATGTGTGCAT
GTGTGTACATGAAGCATGGCAGTGTGTGCACAGGTGTGCAAGGGCACAAGTGTGTGCACATGCGAATGCACACCTGACA
TGCAATGTGTGTTCTGTGCACAGTCGTGTGGGCATTACGTGAGGTGCATGCGTGTGGGTGTGAGTGTGAGTAGCATGTGT
GCACATAACATGTATTGAGGGGTCTCGTGTTCACCCCGCTAGGTCTCAGCACCAGTGCCACTCCTTACAGGATGAGAC
25 GGGGTCCCAGGCCTTGGTGGGCTGAGGCTCTGAAGCTGCAGCCCTGAGGGCATTGTCCCATCTGGGCATCCGCGTCCACT
CCCTCTCCTGTGGGCTTCTGTGTCCACTCCCCCTCTCCTGTGGGCATTTACATCCACTCCACTCCCTCTCTCCTGTGGGC
ATCCGCGTCCACTCCCCCTCTGTGTGGGCATCTGCGTCCACTCCCCCTCTGTGTGGGCATTTGCGTCCACTCCCTCTCCT
GGTTCCTTCTGTCTTGGCCGAGCCTCGGGGGCAGGCAGATGACACAGAGTCTTGACTCGCCAGGGTGGTTGCGAGCTG
CCGGGTGAGGGCCAGGCCGATTCTACTGGGAAGAGGGATAGTTTCTGTCAAATGTTCTCTTTCTTGTTCATCTGA
30 ATGGATGATAAAGCAAAAAGTAAAACTAAATCCAGAGAGGTTTCTACCGTTTCTCACTCTTCTTGGCGACTCTAG

Intron 13 (SEQ ID NO 17)

GTGAGCCGCCACCAAGGGGTGCAGGCCAGCCTCCAGGGACCCTCCGCGCTCTGCTCACCTCTGACCCGGGGCTTCACTT
TGGAACCTCTGGGTTTTAGGGGCAAGGAATGTCTTACGTTTTTCAAGTGGTGTGCTGCTGCTGTGCACAGTCTGTTCGCGTG
35 GCTCTGTGCAAAGCACCTGTCTCCATCTCTGGGTAGTGGTAGGAGCCGGTGTGGCCCCAGGTGTCCCCACTGTGCTGT
GCACTGGCCGTGGGACGTCTATGGAGGCCATCCAGGGCAGCAGGGGCATGGGGTAAAGAGATGTTTATGGGGAGTCTTAG
CAGAGGAGGCTGGGAAGGTGTCTGAACAGTAGATGGGAGATCAGATGCCCGAGGATTTGGGGTCTCAGCAAAGAGGGCC
GAGGTGGGTGCAGGTGAGGGTCGTGGCCCCACCCCGGAAGGTGCAGCAGAGCTGTGGCTCCCCACAGCCCGGCCA
GCACCTGTGCTCTGGGCATGGCTGTGCTCCTGGAACGTTCCCTGTCTGGCTGGTCAGGGGGTGGCCCTGCCAAGAATCG
40 ACAACTTTATCAGAGGGAAGGGCCAATCTGTGGAGGCCACAGGGCCAGCTTCTGCTGGAGTCAGGGCAGGTGGTGGC

ACAAGCCTCGGGGCTGTACCAAAGGGCAGTCGGGCACCACAGGCCCGGGCCTCCACCTCAACAGGCCTCCCAGGCACTG
GGAGCTGAATGCCAGGAGGCCGAAGCCCTCGCCCATGAGGGCTGAGAAGGAGTGTGAGCATTGTGTACCCAGGGCCG
AGGCTGCGCGAATTACCGTGACACTTGATGTGAAATGAGGTCGTCGTCTATCGTGAAACCCAGCAAGGGCTCACGGGA
GAGTTTTCCATTACAAGGTCGTACCATGAAAATGGTTTTTAACCGAGTGCTTGCGCCTTCATGCTCTGGCAGGGAGGGC
5 AGAGCCACAGCTGCATGTTACCGCCTTTGCACCAGCTCCAGAGGCTTGGGACCAGGCTGTCTCAGTTCAGGGTGCGTCC
GGCTCAGACCGCCCTCCTCTGCTCTCTCTGCTCAAATCTTCCCTCGTTTGATCTCCTGACGCGTGCTGGG
CCCTCGTGCAAGCTGCTTGACTCCTTTCCGGAACCCCTTGGGGTGTGCTGGATACAGGTGCCACTGAGGACTGGAGGTGT
CTGACACTGTGGTTGACCCAGGTCAGCTGGCGTGCTTGGGGCCTCCTTGGGCCATGATGAGGTGAGAGGAGTTTTCC
CAGGTGAAAACCTCTGGGAACTCCAGGGCCATGTGACCTGCCACCTGCTCCTCCCATATTAGCTCAGTCTTGTCTCTC
10 ATTTCCCAACCAGGGTCTCTAGCTCCGAGGAGCTCCCGTAGAGGGCTGGGCTCAGGCGAGGGCGGCTGAGTTTCCCCAC
CCATGTGGGGACCTTGGGTAGTCGCTTGATTGGGTAGCCCTGAGGAGGCCGAGATGCGATGGGCCACGGGCCGTTTCCA
AACACAGAGTCAGGCACGTGGAAGGCCAGGAATCCCTTCCCTGAGGCAGGAGTGGGAGAACGGAGAGCTGGGCCCCG
ATTTACGGCAGCCAGGCTGCAGTGGGCGAGGCTGTGGTGGTCCACGTGGCGCTGGGGCGGGGTCTGATTCAAATCCGC
TGGGGCTCGGCCTTCTGGCCCGTGCTGGCCGCGCTCCACAGGGCTTGGGGTGACGCCCGACCTCTAGCAGGTGGC
15 TATTTCTCCCTTTGGAAGAGAGCCCTCACCATGCTAGGTGTTTCCCTCCTGGGTGAGGAGCGTGCCGTGTGGCAACC
CCGGGACCTTAGGCTTATTTATTTGTTTAAACATTCTGGGCCTGGCTTCCGTTGTTGCTAAATGGGGAAAAGACATCC
CACCTCAGCAGAGTTACTGAGAGGCTGAAACCGGGGTCTGGCTTGACTGGTGTGATCTCAGGTCAATCCAGAAGTGGCT
CAGGAAGTCAGTGAGACCAGGTACATGGGGGCTCAGGCAGTGGGTGAGATGAGGTACACGGGGGCTCAGGCAGTGGGT
GAGGCCAGGTACATGGGGGCTCAGGCAGTGGGTGAGATGAGGTACACGGGGGCTCAGGCAGAGGTCAGACCAGGTAC
20 ACGGGGCTCTGATCACACGCACATATGAGCACATGTGCACATGTGCTGTTTCATGGTAGCCAGGTCTGTGCACACCTGC
CCCAAAGTCCCAGGAAGCTGAGAGGCCAAAGATGGAGGTGACAGGGCTGGCGCGGTGGCTCACACCTGTAGTCCCAGCA
CTTTGGGAGGCCGAGGCGAGAGGATCCCTTGAGCCAGGAGTTTAAGACCAGCCTGAGCAACATAGTAGAACCCCATCTC
TATGAAAAATAAAACAAAAATTAGCTGAACATGGTGGTGTGCGCCTGTAGTTCCAATACTTGGGAGGCTGAAGTGGGAG
GATCACTTGAGCCAGGAGGTGGAAGCTGCAGTGAGCTGAGATTGCACCACTGTACTGCAGCCTGGGTGACAGAGTGAGA
25 GCCCATCTCAACAACAACAAGAAGACTGACAAATGCAGTTTCTTGGAAAGAAACATTTAGTAGGAACTTAACCTACACA
CAGAAGCCAAGTCGGTGTCTCGGTGTGAGTGAGATGAGATGATGGGTCTCACACCATCACCCAGACCCAGGGTTTATG
CACCAAGGGGCGGGTGGCTCAGAAGGGATGCGCAGGACGTTGATATACGATGACATCAAGTTGTCTGACGAAGGGCAG
GATTTCATGATAAGTACCTGCTGGTACACAAGGAACAATGGATAAACTGGAACCTTAGAGGCCTTCCCGAACAGGGGCT
AATCAGAAGCCAGCATGGGGGCTGGCATCCAGGATGGAGCTGCTTCAGCCTCCACATGCGTGTTCATACAGATGGTGCA
30 CAGAAACGAGGTGTACCTGTGCACACACAGACACGAGCTACTCGCACACACAAGCACACACAGACATGCATGCATGC
ATCCGTGTGTGTGCACCTGTGCCATGAGGAACCCATGCATGTGCATTATGCACGCACACAGGCACCGGTGGGCCCAT
GCCACACCCACGAGCACCCTCTGATTAGGAGGCCTTTCTCTGACGCTGTCCGCCATCCTCTCAG

Intron 14 (SEQ ID NO 18)

35 GTATGTGCAGGTGCCTGGCCTCAGTGGCAGCAGTGCCTGCCTGCTGGTGTAGTGTGTGAGGAGACTGAGTGAATCTGGG
CTTAGGAAGTTCTTACCCCTTTTCGCATCAGGAAGTGGTTTAACCCAACCACTGTGAGGCTCGTCTGCCCCCCTCTCGT
GGGGTGAGCAGAGCACCTGATGGAAGGACAGGAGCTGTCTGGGAGCTGCCATCCTTCCACCTTGCTCTGCTGGGGAA
GCGCTGGGGGCTGGTCTCTCTGTTTGGCCATGGTGGGATTGGGGGGCTGGCCTCTCCTGTTTGGCCCTGTGGTGG
GATTGGGCTGTCTCCCGTCCATGGCACTTAGGGCCCTTGTGCAAAACCCAGGCCAAGGGCTTAGGAGGAGGCCAGGCCAG
40 GCTACCCACCCCTCTCAGGAGCAGAGGCCGCTATCACCACGACAGAGCCCGCGCGCTCTCTGCTTCCAGTCAACC

TCCTCTGCCCCCTGGACACTTTGTCCAGCATCAGGGAGGTTTCTGATCCGTCTGAAATTCAAGCCATGTCGAACCTGCGGT
CCTGAGCTTAACAGCTTCTACTTTCTGTCTTTCTGTGTTGTGGAAATTTACCTGGAGAAGCCGAAGAAAACATTTCTG
TCGTGACTCCTGCGGTGCTTGGGTGCGGACAGCCAGAGATGGAGCCACCCCGCAGACCGTCGGGTGTGGGCAGCTTTCCG
GTGTCTCCTGGGAGGGGAGCTGGGCTGGGCTGTGACTCCTCAGCCTCTGTTTCCCCCAG

5

Intron 15 (WEQ ID NO 19)

GCAAGTGTGGGTGGAGGCCAGTGCAGGGCCCCACCTGCCAGGGGTATCCTTGAACGCCCTGTGTGGGGCGAGCAGCCTC
AGATGTGTGCTGAAGTGCAGACGCCCGGGCCTGACCCTGGGGCCTGGAGCCACGCTGGCAGCCCTATGTGATTAAACG
CTGGTGTCCCCAGGCCACGGAGCCTGGCAGGGTCCCCAACTTCTTGAACCCCTGCTTCCCATCTCAGGGGCGATGGCTCC
CCACGCTTGGGAGCCTTCTGACCCCTGACCTGTGTCTCTCACAGCCTCTTCCCTGGCTGCTGCCCTGAGCTCCTGGGGT
CCTGAGCAAGTTCTCTCCCGCCCCCGCCTCCAGCGTCACTGGGCTGCCTGTCTGCTCGCCCGGTGGAGGGGTGTCTG
TCCCTTCACTGAGGTTCCACACAGCCAGGGCCACGAGGTGCAGGCCCTGCCTGCCCGGCCACCCACACGTCCTAGGAGGG
TTGGAGGATGCCACCTCTGGCCTCTTCTGGAACGGAGTCTGATTTGGCCCCGAG

10

3'-untranskribierter Bereich (SEQ ID NO 20)

ATCTCATGTTTGAATCCTAATGTGCACTGCATAGACACCACTGTATGCAATTACAGAAGCCTGTGAGTGAACGGGGTGGT
GGTCAGTGCAGGGCCCATGGCCTGGCTGTGCATTACGGAAGTCTATGAGTGAATGGGGTTGTGGTCACTGCGGGCCCATG
GCCTGGCTGGGCCTGGGAGGTTTCTGATGTGTGAGGCAGGAGGGGAAGGAGGGTAGGGGATAGACAGTGGGAGCCCCCA
CCCTGGAAGACATAACAGTAAGTCCAGGCCGAAGGGCAGCAGGGATGTGAGGGGCCAGCTTGGGCGCGGGGATGATG
GAGGGCCTGGCCAGGGTGGCAGGGATGATGGGGCCCCAGCTGGGGTGGCAGGGGTGATGGGGGGGCTGGTCTGGGTGG
CGGGGAAGATGGGAAGCCTGGCTGGGCCCCCTCCTCCCTGCCTCCACCTGCAGCCGTGGATCCGGATGTGCTTCCCT
GGTGACATCCTCTGGGCCATCAGCTTTCATGGAGTGGGGGGCAGGGGCATGACACCATCCTGTATAAAATCCAGGATT
CCTCCTCCTGAACGCCCACTCAGGTTGAAAGTCAATTCGCCCTCTGGCCATTCTCTAAGAGTAGACCAGGATTCTG
ATCTCTGAAGGGTGGGTAGGGTGGGGCAGTGGAGGGTGTGGACACAGGAGGCTTCAGGGTGGGGCTGGTATGCTCTCTC
ATCCTCTTATCATCTCCAGTCTCATCTCTCATCTTATCATCTCCAGTCTCATCTGCTTCTCTTATCTCCAGT
CTCATCTGTCTCTTACCATCTCCAGTCTCATCTTATCCTTATCTCCTAGTCTCATCCAGACTTACCTCCCA
GGGCGGGTGCCAGGCTCGCAGTGGAGCTGGACATACGTCCTTCTCAGGCAGAGGAAGTGAAGGATTGCAGAGAACAG
GAGGGCGGGCTCAGAGGGACGCAGTCTTGGGTGAAGAAACAGCCCTCCTCAGAAGTTGGCTTGGGCCACACGAACCG
AGGGCCCTGCGTGAGTGGCTCCAGAGCCTTCCAGCAGTCCCTGGTGGGGCCTTATGGTATGGCCGGGTCTACTGAGTG
CACCTTGGACAGGGCTTCTGGTTTGTAGTGCAGCCCGACGTGCCTGGTGTGGGGTGGGGCTTATGGCCACTGGATATG
GCGTCATTTATTGCTGCTGCTTCAGAGAATGTCTGAGTGACCGAGCCTAATGTGTATGGTGGGCCAAGTCCACAGACTG
TGTGTAATGCACTCTGGTGCCTGGAGCCCCGTATAGGAGCTGTGAGGAAGGAGGGGCTCTTGGCAGCCGGCCTGGGG
GCGCCTTTGCCCTGCAAACTGGAAGGGAGCGGCCCGGGCGCGCTGGGCGGACGACCTCAAGTGAGAGGTTGGACAGAAC
AGGGCGGGGACTTCCAGGAGCAGAGGCCGCTGCTCAGGCACACCTGGGTTTGAATCACAGACCAACaGGTCAGGCCATT
GTTACAGCTATCCATCTTCTACAAAGCTCCAGATTCTGTCTTCTCCGGGTGTTTTTGTGAAATTTTACTCAGGATTACT
TATATTTTTTGTCTAAAGTATTAGACCCCTAAAAAAGGTATTTGCTTTGATATGGCTTAACCTCACTAAGCACCTACTTTAT
TTGTCTGTTTTATTATTATTATTATTATTATTAGAGATGGTGTCTACTCTGTCAACCAGGTTGTTAGTGCACTGGCAC
AGTCATGGCTCGCTGTAGCCGCAACCCCGAGCTCAAGTGATCCTCCGGCCTCAGCTTCCAGAGTGTCTGGGATTACAG
GTGTGAGCCACTGCCCTTGCCTGGCACTTTAAAAAACCACTATGTAAGGTCAGGTCCAGTGGCTTCCACACCTGTCTATCC
CAGTAGTTTGGGAAGCCGAGGCAGAAGGATTGTCTGAGGCCAGGAGTTTGAGACCAGCATGGGTAACATAGGGAGACCCC

20

25

30

35

40

ATCTCTACAAAAATGCAAAAAGTTATCCGGGCGTGGGGTCCAGCATCTGTAGTCCCAGCTGCTCGGGAGGCTGAGTGGG
AGGATCGCTTGAGCCCGGAGGTCTGGCTGCAGTGAGCTGTGATTGTACCATCGCACTCCAGCCTGGGCAACAGAGTGA
GACCTGTCTCAAAAAAAAAAAAAAAAAAGAAGGAGAAGGAGAAGAGAAGAAGGAAGAAGGAAAGAGAAGAAGAAG
5 GAAGAAGGAAGAAAGAAGGAGAAGGAGGCTGTAGGTGCTAGGTAGACTGTCAAATCTCAGAGCAAAATGAAAATAACA
AAGTTTTAAAGGGAAGAAAAACCCAGCTCTTTGGACTTCCTTAGGCCTGAACTTCATCTCAAGCAGCTTCCTTCCACA
GACAAGCGTGTATGGAGCGAGTGAGTTCAAAGCAGAAAGGGAGGAGAAGCAGGCAAGGGTGGAGGCTGTGGGTGACACCA
GCCAGGACCCCTGAAAGGGAGTGGTTGTTTTCTGCCTCAGCCCCAGCTCCTGCCGGTCTGCACTGCTGTAACCGTC
GATGTTGGTGCCAGGTGCCACCTGGGAAGGATGCTGTGCAGGGGGCTTGCCAACTTTGGTGGGTTTCAGAAGCCCCAG
GCCTTGTGGCAGGCACAATTACAGCCCTCCCAAGATGCCACGTCCTTCTCCTGGAACCTGTGAATGTGTACCCG
10 CAAGGCAGAGGCTGGTGAAGGCTGCAGGTGGAATCACGGCTGCCAGTCAGCCGATCTTAAGGTCATCCTGGATTATCTGG
TGGGCTGATATGGCCACAAGGGTCCCTAGAAGTGAGAGAGGGAGGCAGGGGAGAGTCAGAGAGGGGACGTGAGAAGGAC
CACTGGCCACTGCTGGCTTTGAGATGGAGAGGGGGTCCCCAGCCAAGGAATGGGGGCAGCCGCTCCATGCTGGAAAAGC
AAGCAATCCTCCCGGTCCTGAGGGCACAGGCCCTGCCACGCCCTCGATTTAGGCCAGTGGGACCTGTTTCAGCTTTC
CGGCCTCCAGAGCTGTAAGATGATGCGTTTGTGTTAGCCACTAAGCTGCAGTGATTGTCACAGCAGCAAAATGGAATAG
15 CAGTACAGGGAATGAATACAGGGACAGTTCTCAGAGTGACTCTCAGCCCACCCCTGGG

Die Charakterisierung der Exons zeigte interessanterweise, daß die in unserer Patentanmeldung PCT/EP/98/03469 beschriebenen, funktionell wichtigen hTC-Protein-Domänen auf separaten Exons angeordnet sind. Das Telomerase-charakteristische T-Motiv befindet sich auf Exon 3. Die für die katalytische Funktion der Telomerase wichtigen RT (Reverse-Transkriptase)-Motive 1-7 liegen auf folgenden Exons: RT Motiv 1 und 2 auf Exon 4, RT Motiv 4 auf Exon 9, RT Motiv 5 auf Exon 10, RT Motiv 6 und 7 auf Exon 11. RT Motiv 3 liegt verteilt auf Exon 5 und 6 vor (s. Fig. 8).

10

Die Aufklärung der Exon-Intron-Struktur des hTC-Gens zeigt auch, daß die in unserer Patentanmeldung PCT/EP/98/03469 beschriebenen vier Deletions- bzw. Insertions-Varianten der hTC-cDNA ebenso wie drei weitere, in der Literatur (Kilian et al., 1997) beschriebene hTC-Insertions-Varianten höchstwahrscheinlich alternative Splice-Produkte darstellen. Wie in Fig. 8 gezeigt, lassen sich die Splice Varianten in zwei Gruppen einteilen: Deletionsvarianten und Insertionsvarianten.

15

Den hTC-Varianten der Deletionsgruppe fehlen spezifische Sequenzabschnitte. Die 36 bp in frame Deletion in Variante DEL1 resultiert höchstwahrscheinlich aus der Benutzung einer alternativen 3'-Splice Akzeptorsequenz in Exon 6, wodurch ein Teil des RT Motivs 3 verlorenggeht. In Variante DEL2 werden die normalen 5'-Splice Donor- und 3'-Splice-Akzeptor Sequenzen von Intron 6, 7 und 8 nicht benutzt. Stattdessen wird Exon 6 direkt an Exon 9 fusioniert, wodurch eine Verschiebung des offenen Leserahmens entsteht und in Exon 10 ein Stopcodon auftritt. Variante Del3 stellt eine Kombination aus Variante 1 und 2 dar.

20

25

Die Gruppe der Insertions-Varianten zeichnet sich durch die Insertion von Intronsequenzen aus, die zu vorzeitigen Translationsstop führen. Anstelle der normalerweise benutzten 5'-Splice Donorsequenz von Intron 5 wird eine alternative, 3'-lokalisierte Splicestelle in Variante INS1 benutzt, wodurch eine Insertion der ersten 38 bp aus Intron 4 zwischen Exon 4 und Exon 5 entsteht. Ebenso resultiert die

30

Insertion eines Intron 11-Sequenzbereichs in Variante INS2 aus der Benutzung einer alternativen 5'-Splice Donorsequenz in Intron 11. Da diese Variante in der Literatur (Kilian et al., 1997) nur unzureichend beschrieben wurde, läßt sich die genaue alternative 5'-Splice Donorsequenz dieser Variante nicht bestimmen. Die Insertion von Intron 14 Sequenzen zwischen Exon 14 und Exon 15 in Variante INS3 entsteht durch die Benutzung von einer alternativen 3'-Splice Akzeptorsequenz, wodurch der 3'-Teil von Intron 14 nicht gesplitt wird.

Die in unserer Patentanmeldung PCT/EP/98/03469 beschriebene hTC-Variante INS4 (Variante 4) zeichnet sich durch den Ersatz von Exon 15 und dem 5'-Teilbereich von Exon 16 durch die ersten 600 bp des Introns 14 aus. Diese Variante ist auf den Gebrauch einer alternativer internen 5'-Splice Donorsequenz in Intron 14 und einer alternativen 3'-Splice Akzeptorsequenz in Exon 16 zurückzuführen, woraus ein veränderter C-Terminus resultiert.

Die *in vivo*-Generation wahrscheinlich nicht-funktioneller hTC-Proteinvarianten, die mit der Funktion des vollständigen hTC-Proteins interferieren könnten, stellt zusätzlich zur Transkriptionsregulation einen möglichen Mechanismus dar, um die hTC-Proteinfunktion zu kontrollieren. Bis heute ist die Funktion der hTC-Splicevarianten nicht bekannt. Obwohl die meisten dieser Varianten vermutlich für Proteine ohne Reverse-Transkriptase-Aktivität kodieren, könnten sie dennoch eine entscheidende Rolle als transdominant-negative Telomerase-Regulatoren spielen, indem sie z.B. um die Interaktion mit wichtigen Bindungspartnern kompetieren.

Die Suche nach möglichen Transkriptionsfaktorbindungsstellen wurde mit dem „Find Pattern“-Algorithmus aus dem „GCG Sequenz Analysis“ Programmpaket der „Genetics Computer Group“ (Madison, USA) durchgeführt. Dadurch wurden verschiedene potentielle Bindungsstellen für Transkriptionsfaktoren in der Nukleotidsequenz von Intron 2 identifiziert, die in der Tab. 2 aufgelistet sind. Darüberhinaus wurde im Intron 1 eine Sp1-Bindungsstelle (Pos. 43) und im 5'-

untranslatiertem Bereich eine c-Myc-Bindungsstelle (cDNA-Position 29-34, vergl. Fig. 6) gefunden.

Beispiel 6

5

Um den oder die Startpunkt(e) der hTC-Transkription in HL 60 Zellen zu ermitteln, wurde das 5'-Ende der hTC-mRNA durch Primer-Extension-Analyse bestimmt.

Es wurden 2 µg PolyA⁺-RNA aus HL-60-Zellen für 10 min bei 65°C denaturiert. Zur
10 Primeranlagerung wurden 1 µl RNasin (30-40 U/ml) und 0,3-1 pmol radioaktiv
markierter Primer (5'-GTTAAGTTGTAGCTTACACTGGTTCTC 3'; 2,5-8x10⁵
cpm) zugegeben und für 30 min bei 37°C in einem Gesamtvolumen von 20 µl
inkubiert. Nach Zugabe von 10 µl 5xReverse Transkriptase-Puffer (Fa. Gibco-BRL),
2 µl 10 mM dNTPs, 2 µl RNasin (s.o.), 5 µl 0,1 M DTT (Fa. Gibco-BRL) 2 µl
15 ThermoScript RT (15 U/µl; Fa. Gibco-BRL) und 9 µl DEPC-behandeltes Wasser
erfolgte die Primer-Verlängerung in einem Gesamtvolumen für 1 h bei 58°C. Die
Reaktion wurde durch 4 µl 0,5 M EDTA, pH 8,0, gestoppt und die RNA nach
Zugabe von 1 µl RNaseA (10 mg/ml) für 30 min bei 37°C abgebaut. Hierauf wurden
2,5 µg gescherte Kalbsthymus-DNA und 100 µl TE addiert und einmal mit 150 µl
20 Phenol/Chloroform (1:1) extrahiert. Die DNA wurde unter Zusatz von 15 µl 3 M Na-
Acetat und 450 µl Ethanol für 45 min bei -70°C gefällt und anschließend für 15 min
bei 14000 Upm abzentrifugiert. Das Präzipitat wurde einmal mit 70 %igem Ethanol
gewaschen, luftgetrocknet und in 8 µl Sequenzierungs-Stopplösung gelöst. Nach 5
min Denaturierung bei 80°C wurden die Proben auf ein 6 %iges Polyacrylamidgel
25 aufgetragen und elektrophoretisch (Ausubel et al., 1987) aufgetrennt (Fig. 5).

Hierbei wurde eine Haupt-Transkriptionsstartstelle identifiziert, die 1767 bp 5' vom
ATG-Startcodon der hTC-cDNA Sequenz lokalisiert ist (Nukleotidposition 3346 in
Fig. 4). Die Nukleotidsequenz um diesen Haupttranskriptionsstart (TTA_nTTGT)
30 repräsentiert darüberhinaus ein Initiator-Element (Inr), das in 6 von 7 Nukleotiden

mit dem Konsensusmotiv (PyPyA_nNa/tPyPy) (Smale, 1997) eines Initiator-Elementes übereinstimmt.

In unmittelbarer Nähe des experimentell identifizierten Haupt-Transkriptionsstartes konnte keine eindeutige TATA-Box identifiziert werden, so daß der hTC-Promoter wahrscheinlich in die Familie der TATA-losten Promotoren (Smale, 1997) einzuordnen ist. Allerdings wurde durch Bioinformatik Analyse eine potentielle TATA-Box von Nukleotidposition 1306 bis 1311 (Fig. 4) gefunden. Die zusätzlich um den Haupt-Transkriptionsstart beobachteten Neben-Transkriptionsstarts wurden auch bei anderen TATA-losten Promotoren beschrieben (Geng and Johnson, 1993), wie z.B. in den stark regulierten Promotoren einiger Zellzyklusgene (Wick *et al.*, 1995).

Beispiel 7

Zusätzlich zu dem in Beispiel 6 beschriebenen, in HL60 Zellen identifizierten Startpunkt des hTC Transkriptes, wurde ein weiterer Transkriptionsstartbereich in HL60 Zellen identifiziert. Anhand von RT-PCR-Analysen wurde die Region des Transkriptionsstarts des hTC-Gens in HL60 Zellen auf die bp -60 bis -105 eingegrenzt.

Unter Einsatz von 0,4 µg Poly A-RNA aus HL60 Zellen (Clontech) und dem genspezifischen Primer GSP13 (5'-CCTCCAAAGAGGTGGCTTCTTCGGC-3', cDNA-Position 920-897) wurde hierfür die cDNA mit Hilfe des „First Strand cDNA-Synthesis Kit“ (Clontech) nach Angaben der Hersteller synthetisiert. In einem Endvolumen von 50 µl wurden 1 µl cDNA mit 10 pmol dNTP-Mix versetzt und in 1xPCR-Reaktionspuffer F (PCR-Optimizer Kit der Fa. InVitrogen) und einem Unit Platinum-Taq-DNA Polymerase (Fa. Gibco/BRL) eine PCR-Reaktion durchgeführt. Als Primer wurden jeweils 10 pmol der nachfolgend definierten 5'- und 3'-Primer zugefügt. Die PCR wurde in 3 Schritten durchgeführt. An eine zweiminütige Denaturierung bei 94°C schlossen sich 36 PCR-Zyklen an, in denen die DNA zunächst für 45 sec bei 94°C denaturiert wurde und anschließend für 5 min bei 68°C die Primer

angelagert und die DNA-Kette verlängert wurde. Zum Abschluß folgte für 10 min eine Kettenverlängerung bei 68°C. Insgesamt wurden sechs verschiedene 5'-PCR Primer (Primer HTRT5B: 5'-CGCAGCCACTACCGCGAGGTGC-3', cDNA-Position 105 bis 126; Primer C5S: 5'-CTGCGTCCTGCTGCGCACGTGGGAAGC-3', 5'-flankierende Region -49 bis -23; Primer PRO-TEST1: 5'-CTCGCGGCGCGAGTTTCAGGCAG-3', 5'-flankierende Region -74 bis -52; Primer PRO-TEST2: 5'-CCAGCCCCTCCCCTTCCTTTCC-3', 5'-flankierende Region -112 bis -91; Primer PRO-TEST4: 5'-CCAGCTCCGCCTCCTCCGCGC-3', 5'-flankierende Region -191 - -171; Primer RP-3A: 5'-CTAGGCCGATTGACCTCTCTCC-3', 5'-flankierende Region -427 bis -405) mit dem 3'-PCR Primer C5Rück (5'-GTCCCAGGGCAGCACACCAG-3', cDNA-Position 245 bis 225) kombiniert. Als Kontrolle wurde zusätzlich zu den Oligo-dT- und GSP13-geprinten cDNAs auch genomische DNA für die PCR eingesetzt. Wie in Fig. 9 gezeigt, wurde nur mit den Primerkombinationen HTRT5B-C5Rück, C5S-C5Rück und PRO-TEST1-C5Rück ein PCR-Produkt erhalten, was darauf hinweist, daß der Startpunkt der hTC-Transkription in der Region zwischen bp-60 und bp-105 liegt.

Beispiel 8

In der ca. 11,2 kb isolierten 5'-flankierenden Region des hTC-Gens befinden sich mehrere extrem GC-reiche Bereiche, sog. CpG Islands. Ein CpG Islands mit einem GC-Gehalt von > 70 % reicht von bp - 1214 bis in Intron 2. Zwei weitere GC-reiche Bereiche mit einem GC-Gehalt von > 60 % reichen von bp -3872 bis bp -3113 bzw. bp -5363 bis bp -3941. Die Lage der CpG Islands ist in der Fig. 11 graphisch dargestellt.

Die Suche nach möglichen Transkriptionsfaktorbindungsstellen wurde mit dem „Find Pattern“-Algorithmus aus dem „GCG Sequenz Analysis“ Programmpaket der „Genetics Computer Group“ (Madison, USA) durchgeführt. Dadurch wurden verschiedene potentielle Bindungsstellen in der Region bis -900 bp upstream vom

Translations-Startcodon ATG indentifiziert: fünf Sp1-Bindungsstellen, eine c-Myc-Bindungsstelle, eine CCAC-Box (Fig. 10). Zusätzlich wurden eine CCAAT-Box und eine zweite c-Myc-Bindungsstelle an den Positionen -1788 bzw. -3995 der 5'-flankierenden Region gefunden.

5

Beispiel 9

Um die Aktivität des hTC-Promotors zu analysieren, wurden durch PCR-Amplifikation vier verschieden lange hTC-Promotorsequenzabschnitte generiert und 5' vor das Reportergen Luziferase in den Vektor pGL2 der Fa. Promega kloniert. Als DNA-Quelle für die PCR-Amplifikation wurde das aus dem Phagenklon P12 subklonierte, 8,5 kb große SacI-Fragment gewählt. In einem Endvolumen von 50 µl wurden 35 ng dieser DNA mit 10 pmol dNTP-Mix versetzt und in 1xPCR-Reaktionspuffer (PCR-Optimizer Kit der Fa. InVitrogen) und einem Unit Platinum-Taq-DNA Polymerase (Fa. Gibco/BRL) eine PCR-Reaktion durchgeführt. Als Primer wurden jeweils 20 pmol der nachfolgend definierten 5'- und 3'-Primer zugefügt. Die PCR wurde in 3 Schritten durchgeführt. An eine zweiminütige Denaturierung bei 94°C schlossen sich 30 PCR-Zyklen an, in denen die DNA zunächst für 45 sec bei 94°C denaturiert wurde und anschließend für 5 min bei 68°C die Primer angelagert und die DNA-Kette verlängert wurde. Zum Abschluß folgte für 10 min eine Kettenverlängerung bei 68°C. Als 3'-PCR-Primer wurde jeweils der Primer PK-3A (5'-GCAAGCTTGACGCAGCGCTGCCTGAAACTCG-3', Position -43 bis -65) gewählt, der einen Sequenzbereich 42 bp upstream vom START-Codon ATG erkennt. Durch Kombination des PK-3A-Primers mit dem 5'-PCR-Primer PK-5B (5'-CCAGATCTCTGGAACACAGAGTGGCAGTTTCC-3', Position -4093 bis -4070) wurde ein 4051 bp großes Promotor-Fragment amplifiziert (NPK8). Die Kombination des Primerpaares PK-3A und PK-5C (5'-CCAGATCTGCATGAAGTGTGTGGGGATTTGCAG-3', Position -3120 bis -3096) führte zur Amplifikation eines 3078 bp großen Promotorfragmentes (NPK15). Ein 2068 bp großes Promotorfragment wurde durch die Verwendung der Primerkombination PK-3A und PK-5D (5'-

GGAGATCTGATCTTGGCTTACTGCAGCCTCTG-3', Position -2110 bis -2087) amplifiziert (NPK22). Der Einsatz der Primerkombination PK-3A und PK-5E (5'-GGAGATCTGTCTGGATTCCTGGGAAGTCCTCA-3', Position -1125 bis -1102) führte schließlich zur Amplifikation eines 1083 bp großen Promotorfragmentes (NPK27). Der PK-3A Primer enthält eine HindIII Erkennungssequenz. Die
5 verschiedenen 5'-Primer enthalten eine BglII-Erkennungssequenz.

Die entstandenen PCR-Produkte wurden mit Hilfe des QIA quick spin PCR Purification Kits der Fa. Qiagen nach Angaben der Hersteller aufgereinigt und
10 anschließend mit den Restriktionsenzymen BglII und HindIII verdaut. Mit den gleichen Restriktionsenzymen wurde der pGL2-Promotor-Vektor verdaut und der in diesem Vektor enthaltene SV40-Promotor freigesetzt und abgetrennt. Die PCR-Promotorfragmente wurden in den Vektor ligiert, in kompetente DH5 α -Bakterien der Fa. Gibco/BRL transformiert. Aus transformierten Bakterienklonen wurde DNA für
15 die nachfolgend beschriebenen Promotor-Aktivitäts-Analysen mit Hilfe des Qiagen Plasmid-Kits der Fa. Qiagen isoliert.

Beispiel 10

20 Die Aktivität des hTC-Promotors wurde in transienten Transfektionen in eukaryotischen Zellen analysiert.

Alle Arbeiten mit eukaryotischen Zellen erfolgten an einem sterilen Arbeitsplatz. CHO-K1 und HEK 293 Zellen wurden von der American Type Culture collection
25 bezogen.

CHO-K1 Zellen wurden in DMEM Nut Mix F-12 Zellkulturmedium (Fa. Gibco-BRL, Bestellnummer: 21331-020) mit 0,15 % Streptomycin/Penezillin, 2 mM Glutamin und 10 % FCS (Fa. Gibco-BRL) gehalten.

30

HEK 293 Zellen wurden in DMOD Zellkulturmedium (Fa. Gibco-BRL, Bestellnummer: 41965-039) mit 0,15 % Streptomycin/Penizillin, 2 mM Glutamin und 10 % FCS (Fa. Gibco-BRL) kultiviert.

- 5 CHO-K1 und HEK 293 Zellen wurden in wasssergesättigter Atmosphäre bei 37°C unter Begasung mit 5 % CO₂ kultiviert. Bei konfluentem Zellrasen wurde das Medium abgesaugt, die Zellen mit PBS (100 mM KH₂PO₄, pH 7,2; 150 mM NaCl) gewaschen und durch Zugabe einer Trypsin-EDTA Lösung (Fa. Gibco-BRL) abgelöst. Das Trypsin wurde durch Mediumzugabe inaktiviert und die Zellzahl mit
10 einer Neubauer-Zählkammer ermittelt, um die Zellen in gewünschter Dichte auszu-plattieren.

- Für die Transfektion wurden pro Well jeweils 2×10^5 –HEK 293 Zellen in einer 24-well Zellkulturplatte ausplattiert. Nach 3 Stunden wurde das HEK 293 Medium
15 entfernt. Für die Transfektion wurden bis zu 2,5 µg Plasmid-DNA, 1 µg eines CMV β-Gal Plasmidkonstruktes (Fa. Stratagene, Bestellnummer: 200388), 200 µl serum-freies Medium und 10 µl Transfektionsreagenz (DOTAP der Fa. Boehringer Mannheim) für 15 Minuten bei Raumtemperatur inkubiert und anschließend auf die HEK 293 Zellen gleichmäßig aufgetropft. Nach 3 Stunden wurden 1,5 ml Medium
20 hinzugegeben. Nach 20 Stunden wurde das Medium gewechselt. Nach weiteren 24 Stunden wurden die Zellen zur Bestimmung der Luziferase- und der β-Gal-Aktivität geerntet. Dazu wurden die Zellen im Zellkultur-Lysisreagenz (25 mM Tris [pH 7,8] mit H₃PO₄; 2 mM CDTA; 2 mM DTT; 10% Glycerol; 1% Triton X-100) für 15 Minuten bei Raumtemperatur lysiert. Zwanzig µl dieses Zellysats wurden mit 100 µl
25 Luziferase-Assaypuffer (20 mM Tricin; 1,07 mM (MgCO₃)₄, Mg(OH)₂·5H₂O; 2,67 mM MgSO₄; 0,1 mM EDTA; 33,3 mM DTT; 270 µM Coenzym A; 470 µM Luciferin, 530 µM ATP) gemischt und das durch die Luziferase generierte Licht gemessen.

- 30 Zur Messung der β-Galaktosidaseaktivität wurden gleiche Mengen Zellysats und β-Galaktosidase-Assaypuffer (100 mM Natriumphosphatpuffer pH 7,3; 1 mM MgCl₂;

50 mM β -Merkaptoethanol; 0,665 mg/ml ONPG) für mindestens 30 Minuten bei 37°C oder bis eine leichte Gelbfärbung auftrat, inkubiert. Die Reaktion wurde durch Zugabe von 100 μ l 1 M Na_2CO_3 gestoppt und die Absorption bei 420 nm bestimmt.

- 5 Für die Analyse des hTC-Promotors wurden vier verschieden lange hTC-Promotorsequenzabschnitte 5' vor das Reportergen Luziferase kloniert (vergl. Beispiel 9).

In der Fig. 11 sind die relativen Luziferase Aktivitäten zweier unabhängiger Transfektionen mit den Konstrukten NPK8, NPK15, NPK22 und NPK27 in HEK
10 293 Zellen aufgetragen. Jedes Experiment wurde in Duplikaten durchgeführt. Darüberhinaus wurde die Standardabweichung angegeben. Das Konstrukt NPK 27 zeigt eine 40fach höhere Luziferaseaktivität als die Basalaktivität des promotorlosen Luziferase-Kontrollkonstrutes (pGL2-basic) und eine 2 bis 3fach höhere Aktivität als das SV40 Promotorkontroll-Konstrukt (pGL2PRO). Interessanterweise wurde im
15 Vergleich zu dem Konstrukt NPK27 eine 2 bis 3fach geringere Luziferaseaktivität in mit längeren hTC Promotorkonstrukten (NPK8, NPK15, NPK22) transfizierten Zellen beobachtet. Ähnliche Ergebnisse wurden auch in CHO Zellen beobachtet (Daten nicht gezeigt).

Literaturverzeichnis

- 5 Allsopp, R. C., Vazire, H., Pattersson, C., Goldstein, S., Younglai, E.V., Futcher, A.B., Greider, C.W. und Harley, C.B. (1992). Telomere length predicts replicative capacity of human fibroblasts. Proc. Natl. Acad. Sci. 89, 10114-10118.
- 10 Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., Struhl, K. (1987). Current protocols in molecular biology. Greene Publishing Associates and Wiley-Intersciences, New York.
- Blasco, M. A., Rizen, M., Greider, C. W. und Hanahan, D. (1996). Differential regulation of telomerase activity and telomerase RNA during multistage tumorigenesis. Nature Genetics 12, 200-204.
- 15 Broccoli, D., Young, J. W. und deLange, T. (1995). Telomerase activity in normal and malignant hematopoietic cells. Proc. Natl. Acad. Sci. 92, 9082-9086.
- Counter, C. M., Ailion, A. A., LeFeuvre, C. E., Stewart, N. G. Greider, C.W. Harley, C. B. und Bacchetti S. (1992). Telomere shortening associated with chromosome instability is arrested in immortal cells which express telomerase activity. EMBO J. 11, 1921-1929.
- 20 Feng, J., Funk, W. D., Wang, S.-S., Weinrich, S. L., Ailion, A.A., Chiu, C.-P., Adams, R.R., Chang, E., Allsopp, R.C., Yu, J., Le, S., West, M.D., Harley, C.B., Andrews, W.H., Greider, C.W. und Villeponteau, B. (1995). The RNA component of human telomerase. Science 269, 1236-1241.
- 25 Geng, Y., and Johnson, L.F. (1993). Lack of an initiator element is responsible for multiple transcriptional initiation sites of the TATA less mouse thymidine synthase promoter. Mol. Cell. Biol. 14:4894.
- 30 Goldstein, S. (1990). Replicative senescence: The human fibroblast comes of age. Science 249, 1129-1133.
- 35 Harley, C.B., Futcher, A.B., Greider, C.W., 1990. Telomeres shorten during ageing of human fibroblasts. Nature 345, 458-460.

- Hastie, N. D., Dempster, M., Dunlop, M. G., Thompson, A. M., Green, D.K. und Allshire, R.C. (1990). Telomere reduction in human colorectal carcinoma and with ageing. *Nature* 346, 866-868.
- 5 Hiyama, K., Hirai, Y., Kyoizumi, S., Akiyama, M., Hiyama, E., Piatyszek, M.A., Shay, J.W., Ishioka, S. und Yamakido, M. (1995). Activation of telomerase in human lymphocytes and hematopoietic progenitor cells. *J. Immunol.* 155, 3711-3715.
- 10 Kim, N.W., Piatyszek, M.A., Prowse, K.R., Harley, C. B., West, M.D., Ho, P.L.C., Coviello, G.M., Wright, W.E., Weinrich, S.L. und Shay, J.W. (1994). Specific association of human telomerase activity with immortal cells and cancer. *Science* 266, 2011-2015.
- Latchman, D.S. (1991). Eukaryotic transcription factors. Academic Press Limited, London.
- 15 Lingner, J., Hughes, T.R., Shevchenko, A., Mann, M., Lundblad, V. und Cech T.R. (1997). Reverse transcriptase motifs in the catalytic subunit of telomerase. *Science* 276: 561-567.
- Lundblad, V. und Szostak, J. W. (1989). A mutant with a defect in telomere elongation leads to senescence in yeast. *Cell* 57, 633-643.
- 20 McClintock, B. (1941). The stability of broken ends of chromosomes in *Zea mays*. *Genetics* 26, 234-282.
- 25 Meyne, J., Ratliff, R. L. und Moyzis, R. K. (1989). Conservation of the human telomere sequence (TTAGGG)_n among vertebrates. *Proc. Natl. Acad. Sci.* 86, 7049-7053.
- Olovnikov, A. M. (1973). A theory of marginotomy. *J. Theor. Biol.* 41, 181-190.
- 30 Sandell, L. L. und Zakian, V. A. (1993). Loss of a yeast telomere: Arrest, recovery and chromosome loss. *Cell* 75, 729-739.
- Shapiro, M.B., Senapathy, P., 1987. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucl. Acids Res.* 15, 7155-7174.
- 35 Smale, S.T. and Baltimore, D. (1989). The „initiator“ as a transcription control element. *Cell* 57:103-113.

Smale, S.T. (1997). Transcription initiation from TATA-less promoters within eukaryotic protein-coding genes. *Biochimica et Biophysica Acta* 1351, 73-88.

5 Shay, J. W. (1997). Telomerae and Cancer. Ciba Foundation Meeting: Telomeres and Telomerase. London.

Vaziri, H., Dragowska, W., Allsopp, R. C., Thomas, T. E., Harley, C.B. und Landsdorp, P.M. (1994). Evidence for a mitotic clock in human hematopoietic stem cells: Loss of telomeric DNA with
10 age. *Proc. Natl. Acad. Sci.* 91, 9857-9860.

Wick, M., Härönen, R., Mumberg, D., Bürger, C., Olsen, B.R., Budarf, M.L., Apte, S. S. and Müller, R. (1995). Structure of the human TIMP-3 gene and its cell-cycle-regulated promoter. *Biochemical Journal* 311, 549-554.

15 Zakian, V. A. (1995). Telomeres: Beginning to understand the end. *Science* 270, 1601-1607.

Patentansprüche

1. Regulatorische DNA-Sequenzen für das Gen der humanen katalytischen Telomerase-Untereinheit.
2. DNA-Sequenzen gemäß Anspruch 1, dadurch gekennzeichnet, daß es sich um Intronsequenzen gemäß SEQ ID NO 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 und/oder 20 oder um regulatorisch wirksame Fragmente dieser Sequenzen handelt.
3. DNA-Sequenzen gemäß Anspruch 1, dadurch gekennzeichnet, daß es sich um die 5'-flankierende regulatorische DNA-Sequenz für das Gen der humanen katalytischen Telomerase-Untereinheit gemäß Fig. 10 (SEQ ID NO 3) oder um regulatorisch wirksame Fragmente dieser DNA-Sequenz handelt.
4. Rekombinantes Konstrukt, enthaltend eine DNA-Sequenz gemäß einem der Ansprüche 1 bis 3.
5. Rekombinantes Konstrukt gemäß Anspruch 4, dadurch gekennzeichnet, daß es weiterhin eine oder mehrere DNA-Sequenzen enthält, die für Polypeptide oder Proteine kodieren.
6. Vektor, enthaltend ein rekombinantes Konstrukt gemäß Anspruch 4 oder 5.
7. Verwendung von rekombinanten Konstrukten bzw. Vektoren gemäß einem der Ansprüche 4 bis 6 zur Herstellung von Arzneimitteln.
8. Rekombinante Wirtszellen, enthaltend rekombinante Konstrukte bzw. Vektoren gemäß einem der Ansprüche 4 bis 6.

9. Verfahren zur Identifizierung von Substanzen, die die Promotor-, Silencer- oder Enhanceraktivität der humanen katalytischen Telomerase-Untereinheit beeinflussen, das folgende Schritte umfaßt:
- 5 A. Zugabe einer Kandidatensubstanz zu einer Wirtszelle, enthaltend DNA-Sequenzen gemäß einem der Ansprüche 1 bis 3, funktionell verknüpft mit einem Reportergen,
- B. Messung des Substanzeffektes auf die Reportergenexpression.
- 10
10. Verfahren zur Identifizierung von Faktoren, die spezifisch an die DNA gemäß einem der Ansprüche 1 bis 3 oder an Fragmente davon binden, dadurch gekennzeichnet, daß man eine Expressions-cDNA-Bibliothek mit einer DNA-Sequenz gemäß einem der Ansprüche 1 bis 3 oder Teilfragmenten unterschiedlichster Länge als Sonde screent.
- 15
11. Transgene Tiere, enthaltend rekombinante Konstrukte bzw. Vektoren gemäß Ansprüchen 4 bis 6.
- 20
12. Verfahren zur Detektion Telomerase-assoziiierter Zustände bei einem Patienten, das folgende Schritte umfaßt:
- A. Inkubation eines rekombinanten Konstruktes bzw. Vektors gemäß Ansprüchen 4 bis 6 das bzw. der zusätzlich ein Reportergen enthält
- 25 mit Körperflüssigkeiten oder zellulären Proben,
- B. Detektion der Reportergenaktivität, um einen diagnostischen Wert zu erhalten,

- C. Vergleich des diagnostischen Wertes mit Standardwerten für das Reportergenkonstrukt in standardisierten normalen Zellen oder Körperflüssigkeiten des gleichen Typs wie die Testprobe.

Fig. 1

A

1 2 3 4 5 6 7 8 9 10



B

1 2 3 4 5 6 7 8 9 10

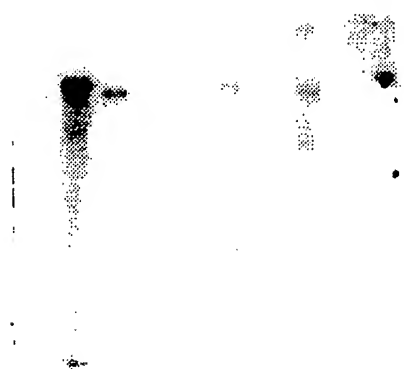


Fig. 2

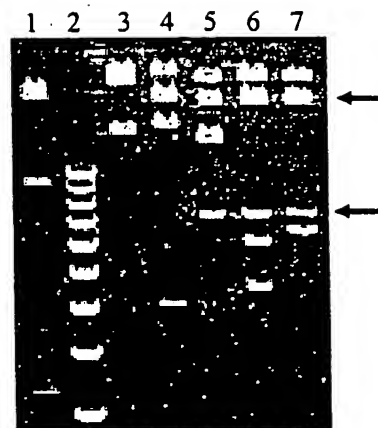


Fig. 3

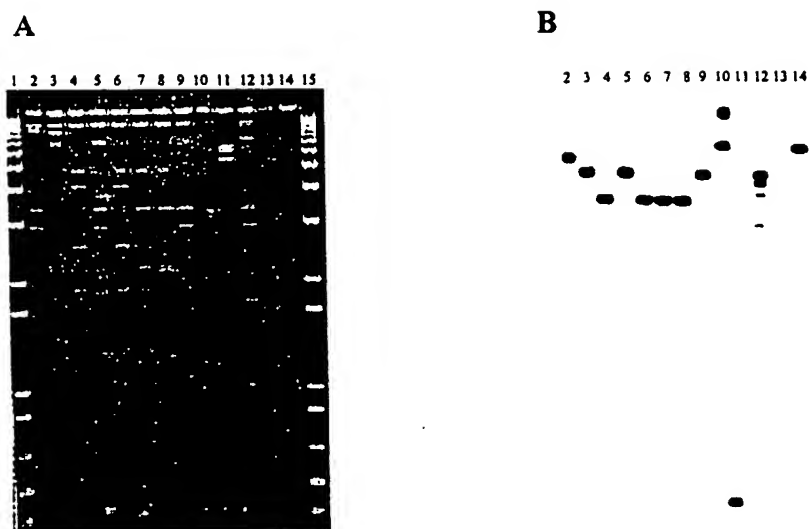


Fig. 4

GAGCTCTGAA	CCGTGGAAAC	GAACATGACC	CTTGCCTGCC	TGCTTCCCTG	GGTGGGTCAA	GGGTAATGAA	70
GTGGTGTGCA	GGAAATGGCC	ATGTAAATTA	CACGACTCTG	CTGATGGGGA	CCGTTCCTTC	CATCATTATT	140
CATCTTCACC	CCCAAGGACT	GAATGATTCC	AGCAACTTCT	TCGGGTGTGA	CAAGCCATGA	CAAAACTCAG	210
TACAAACACC	ACTCTTTTAC	TAGGCCACCA	GAGCAGGGGC	CACACCCTTG	ATATATTAA	AGTCCAGGAG	280
AGATGAGGCT	GCTTTCAGCC	ACCAGGCTGG	GGTGACAACA	GCGGCTGAAC	AGTCTGTTCC	TCTAGACTAG	350
TAGACCCTGG	CAGGCACTCC	CCCAAATTCT	AGGGCCTGGT	TGCTGCTTCC	CGAGGGCGCC	ATCTGCCCTG	420
GAGACTCAGC	CTGGGGTGCC	ACACTGAGGC	CAGCCCTGTC	TCCACACCCT	CCGCCTCCAG	GCCTCAGCTT	490
CTCCAGCAGC	TTCTTAAACC	CTGGGTGGGC	CGTGTTCAG	CGTACTGTC	TCACCTGTCC	CACGTGTCT	560
TGCTCTCAGC	ACGTAGCTCG	CACGGTTCCT	CCTCAGATGG	GGTGTCTGTC	TCCTTCCCCA	ACACTCACAT	630
GCGTTGAAGG	GAGGAGATTG	TGCGCCTCCC	AGACTGGCTC	CTCTGAGCCT	GAACCTGGCT	CGTGGCCCCC	700
GATGCAGGTT	CCTGGCGTCC	GGCTGCACGC	TGACCTCCAT	TTCCAGGCGC	TCCCGCTCTC	CTGTCACTG	770
CCGGGGCCTG	CCGGTGTGTT	CTTCTGTTTC	TGTGCTCCTT	TCCACGTCCA	GCTGCGTGTG	TCTCTGCCCG	840
CTAGGGTCTC	GGGGTTTTTA	TAGGCATAGG	ACGGGGGCGT	GGTGGGCCAG	GGCGCTCTTG	GGAAATGCAA	910
CATTGGGTG	TGAAAGTAGG	AGTGCCTGTC	CTCACCTAGG	TCCACGGGCA	CAGGCCTGGG	GATGGAGCCC	980
CCGCCAGGGA	CCCGCCCTTC	TCTGCCCAG	ACTTTCCTGC	CCCCCTCCCT	CTGGAACACA	GAGTGGCAGT	1050
TTCCACAAGC	ACTAAGCATC	CTCTTCCCAA	AAGACCCAGC	ATTGGCACCC	CTGGACATT	GCCCCACAGC	1120
CCTGGGAATT	CACGTGACTA	CGCACATCAT	GTACACACTC	CCGTCCACGA	CCGACCCCCG	CTGTTTTATT	1190
TTAATAGCTA	CAAAGCAGGG	AAATCCCTGC	TAAATGTCC	TTTAACAAAC	TGGTTAAACA	AACGGGTCCA	1260
TCCGCACGGT	GGACAGTTCC	TCACAGTGA	GAGGAACATG	CCGTTTATAA	AGCCTGCAGG	CATCTCAAGG	1330
GAATTACGCT	GAGTCAAAAC	TGCCACCTCC	ATGGGATACG	TACGCAACAT	GCTCAAAAAG	AAAGAATTTT	1400
ACCCCATGGC	AGGGGAGTGG	TTAGGGGGGT	TAAGGACGGT	GGGGGGCGCA	GCTGGGGGCT	ACTGCACGCA	1470
CCTTTTACTA	AAGCCAGTTT	CCTGGTTCTG	ATGGTATTGG	CTCAGTTATG	GGAGACTAAC	CATAGGGGAG	1540
TGGGGATGGG	GGAAACCCGA	GGCTGTGCCA	TCCTTGCCAT	GCCCGAGTGT	CCTGGGCAGG	ATAATGCTCT	1610
AGAGATGCCC	ACGTCTCTGAT	TCCCCCAAAC	CTGTGGACAG	AACCCGCCCG	GCCCCAGGGC	CTTTGCAAGT	1680
GTGATCTCCG	TGAGGACCTT	GAGGTCTGGG	ATCCTTCGGG	ACTACCTGCA	GGCCCGAAAA	GTAATCCAGG	1750
GGTTCTGGGA	AGAGGCGGGC	AGGAGGGTCA	GAGGGGGGCA	GCCTCAGGAC	GATGGAGGCA	GTCACTCTGA	1820
GGCTGAAAAG	GGAGGGAGGG	CCTCGAGCCC	AGGCCTGCAA	GCGCCTCCAG	AAGCTGGAAA	AAGCGGGGAA	1890
GGGACCCCTC	ACGGAGCCTG	CAGCAGGAAG	GCACGGCTGG	CCCTTAGCCC	ACCAGGGCCC	ATCGTGGACC	1960
TCCGGCCTCC	GTGCCATAGG	AGGGCACTCG	CGCTGCCCTT	CTAGCATGAA	GTGTGTGGGG	ATTTGCAGAA	2030
GCAACAGGAA	ACCCATGCAC	TGTGAATCTA	GGATTATTTT	AAAACAAAGG	TTTACAGAAA	CATCCAAGGA	2100
CAGGGCTGAA	GTGCCCTCCG	GCAAGGGCAG	GGCAGGCACG	AGTGATTTTA	TTTAGCTATT	TTATTTTATT	2170
TACTTACTTT	CTGAGACAGA	GTTATGCTCT	TGTTGCCCAG	GCTGGAGTGC	AGCGGCATGA	CTTGGCTCA	2240
CTGCAACCTC	CGTCTCCTGG	GTTCAAGCAA	TTCTCGTGCC	TCAGCCTCCC	AAGTAGCTGG	GATTTACAGC	2310
GTGCACCACC	ACACCCGGCT	AATTTTGTAT	TTTATGATGA	GATGGGCTTT	CACCATGTTG	GTCAAGCTGA	2380
TCTCAAAATC	CTGACCTCAG	GTGATCCGCC	CACCTCAGCC	TCCCAAAGTG	CTGGGATTAC	AGGCATGAGC	2450
CACCTGCACCT	GGCCTATTTA	ACCAATTTAA	AACCTCCCTG	GGCTCAAGTC	ACACCCACTG	GTAAGGAGTT	2520
CATGGAGTTC	AATTTCCCTT	TTACTCAGGA	GTTACCTCC	TTTGATATTT	TCTGTAATTC	TTCTGAGACT	2590
GGGGATACAC	CGTCTCTTGA	CATATTACCA	GTTTCTGTGA	CCACCTGTTA	TCCCATGGGA	CCCACTGCAG	2660
GGGCAGCTGG	GAGGCTGCAG	GCTTCAGGTC	CCAGTGGGGT	TGCCATCTGC	CAGTAGAATA	CTGATGTAGA	2730
ATCAGGGGCG	AAGTGTGGAC	ACTGTCCTGA	ATCTCAATGT	CTCAGTGTGT	GCTGAAACAT	GTAGAAATTA	2800
AAGTCCATCC	CTCCTACTCT	ACTGGGATTG	AGCCCCCTCC	CTATCCCCCC	CCAGGGGCGC	AGGAGTTCCT	2870
CTCACTCCTG	TGGAGGAAGG	AATGATACTT	TGTTATTTT	CACTGCTGGT	ACTGAATCCA	CTGTTTCATT	2940
TGTTGGTTTG	TTTGTTTTGT	TTTGAGAGGC	GGTTTCACTC	TTGTTGCTCA	GGCTGGAGGG	AGTGCAATGG	3010
CGCGATCTTG	GCTTACTGCA	GCCTCTGCCT	CCCAGGTTCA	AGTGATTCTC	CTGCTTCCCG	CTCCCATTTG	3080
GCTGGGATTA	CAGGCACCCG	CCACCATGCC	CAGCTAATTT	TTTGTATTTT	TAGTAGAGAC	GGGGGTGGGT	3150
GGGGTTACCC	ATGTTGGCCA	GGCTGGTCTC	GAATTTCTGA	CCTCAGATGA	TCCACCTGCC	TCTGCCCTCT	3220
AAAGTGTCTG	GATTACAGGT	GTGAGCCACC	ATGCCACGCT	CAGAAATTAC	TCTGTTTAGA	AACATCTGGG	3290
TCTGAGGTAG	GAAGCTCACC	CCACTCAAGT	GTTGTGGTGT	TTTAAGCCAA	TGATAGTAAT	TTTTTATTGT	3360
TGTTAGAACCA	CTCTTGATGT	TTTACACTGT	GATGACTAAG	ACATCATCAG	CTTTTCAAAG	ACACACTAAC	3430
TGCACCCATA	ATACTGGGGT	GTCTTCTGGG	TATCAGCAAT	CTTCATTGAA	TGCCGGGAGG	CGTTTCTCTG	3500
CCATGCACAT	GGTGTTAATT	ACTCCAGCAT	AATCTTCTGC	TTCCATTCTT	TCTCTTCCCT	CTTTTAAAT	3570
TGTGTTTTCT	ATGTTGGCTT	CTCTGCAGAG	AACCAGTGTA	AGCTACAAC	TAACTTTTGT	TGGAACAAAT	3640
TTTCCAAACC	CCCCCTTTGC	CCTAGTGGCA	GAGACAATTC	ACAAACACAG	CCCTTTAAAA	AGGCTTAGGG	3710
ATCACTAAGG	GGATTCTTAG	AAGAGCGACC	TGTAATCCTA	AGTATTTACA	AGACGAGGCT	AACCTCCAGC	3780
GAGCGTGACA	CCCCAGGGAG	GGTGCGAGGC	CTGTTCAAAT	GCTAGCTCCA	TAAATAAAGC	AATTTCTCTC	3850
GGCAGTTTCT	GAAGTAGGAA	AAGGTTACAT	TTAAGGTTGC	GTTTGTAGC	ATTTCACTGT	TTGCCGACCT	3920
CAGCTACAGC	ATCCCTGCAA	GGCCTCGGGA	GACCCAGAAG	TTTCTCGCCC	CCTTAGATCC	AAACTTGAGC	3990
AACCCGGAGT	CTGGATTCTT	GGGAAGTCTT	CAGCTGTCTT	GCGGTTGTGC	CGGGGCCCCA	GGTCTGGAGG	4060
GGACCAATGG	CCGTGTGGCT	TCTACTGCTG	GGCTGGAAGT	CGGGCTCCTT	AGCTCTGCAG	TCCGAGGCTT	4130
GGAGCCAGGT	GCCTGGACCC	CGAGGCTGCC	CTCCACCCTG	TGCGGGCGGG	ATGTGACCAG	ATGTTGGCCT	4200
CATCTGCCAG	ACAGAGTGCC	GGGGCCACAG	GTCAAGGCCG	TTGTGGCTGG	TGTGAGGCGC	CCGCTGCCCG	4270
GCCACAGGGA	GCGCCTGGCT	CCATTCCCA	CCCTTCTCG	ACGGGACCGC	CCCGGTGGGT	GATTAAACAG	4340
TTTGGGGTGG	TTTGCTCATG	GTGGGGACCC	CTCGCCGCTT	GAGAACCCTG	AAAGAGAAAT	GACGGGCCCTG	4410
TGTCAGGAGG	CCCAAGTCGC	GGGGAAGTGT	TGCAGGGAGG	CACCTCCGGGA	GGTCCCGCGT	GCCCGTCCAG	4480
GGAGCAATGC	GTCTCCGGGT	TCGTCGCCAG	CCGCTCTAC	GCGCTCCTGT	CCTCCCTTTC	ACGTCGGGCA	4550
TTCTGTGTGC	CCGGAGCCCG	ACGCCCGCGG	TCCGGACCTG	GAGGCAGCCC	TGGGTCTCCG	GATCAGGCCA	4620
GCGGCCAAAG	GGTCCCGCCA	CGCACCTGTT	CCAGGGGCTT	CCACATCATG	GCCCTCCTCT	CGGGTACCC	4690

Fig. 4 (Fortsetzung)

```
CACAGCCTAG GCCGATTGGA CCTCTCTCCG CTGGGGCCCT CGCTGGCGTC CCTGCACCCT GGGAGCGCGA 4760
GGCGCGCGCG GCGGGGGAAG CGCGGCCAG ACCCCCGGGT CCGCCCGGAG CAGCTGCGCT GTCGGGGCCA 4830
GGCCGGGCTC CCAGTGGATT CGCGGGCACA GACGCCAGG ACCGCGCTCC CCACGTGGCG GAGGGACTGG 4900
GGACCCGGGC ACCCGTCCTG CCCCTTCACC TTCCAGCTCC GCCTCCTCCG CGCGGACCCC GCCCCGTCCC 4970
GACCCCTCCC GGGTCCCCGG CCCAGCCCCC TCCGGGCCCT CCCAGCCCCC CCCCTTCCTT TCCGCGGGCC 5040
CGCCCTCTCC TCGCGGCGCG AGTTTCAGGC AGCGCTGCGT CCTGCTGCGC ACGTGGGAAG CCCTGGCCCC 5110
GGCCACCCCC GCGATG 5126
```

Fig. 5

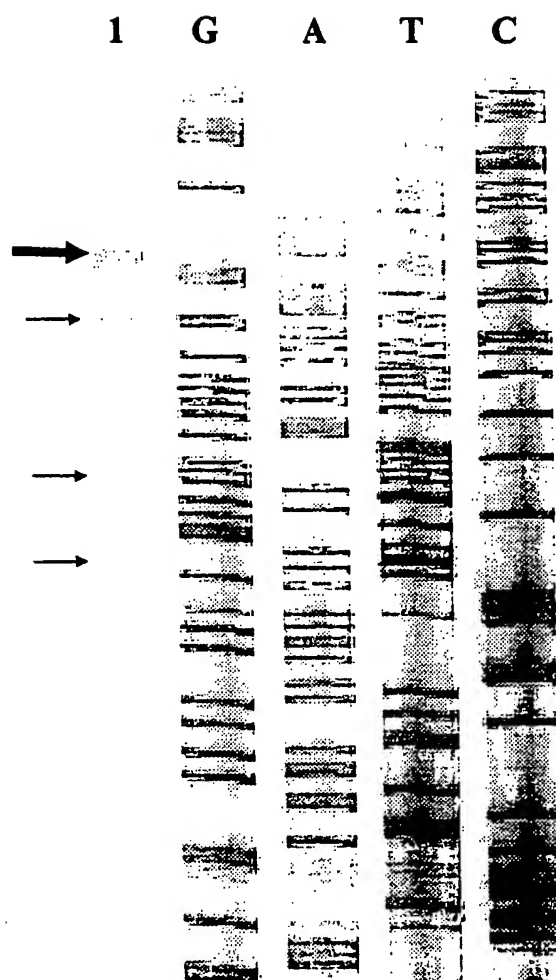


Fig. 6

```

GTTTCAGGCA GCGCTGCGTC CTGCTGCGCA CGTGGGAAGC CCTGGCCCCG GCCACCCCCG CGATGCCCGG 70
CGCTCCCCCG TGCCGAGCCG TGCCTCCCT GCTGCGCAGC CACTACCGCG AGGTGCTGCC GCTGGCCACG 140
TTCGTGCGGC GCCTGGGGCC CCAGGGCTGG CGGCTGGTGC AGCGCGGGGA CCGGCGGGCT TCCGCGCGCG 210
TGTGGGCGCA GTGCTGGTGG TGCCTGCCCT GGGACGCACG GCCGCCCCCG CCGCCCCCTT CCTTCCGCCA 280
GGTGTCTGCG CTGAAGGAGC TGGTGGCCCG AGTGTGTCAG AGGCTGTGCG AGCGCGGGCG GAAGAACGTG 350
CTGGCCCTTCG GCTTCGCGCT GCTGGACGGG GCCCGCGGGG GCGCCCCCGA GGCCTTCACC ACCAGCGTGC 420
GCAGCTACCT GCCCAACACG GTGACCGACG CACTGCGGGG GAGCGGGGCG TGGGGGCTCG TGTGCGCGCG 490
CGTGGGCGAC GACGTGCTGG TTCACCTGCT GGCACGCTGC GCGCTCTTTG TGTGGTGGG TCCAGCTGCG 560
GCCTACCAGG TGTGCGGGCC GCCGCTGTAC CAGCTCGGCG CTGCCACTCA GCGCGGGCCC CCGCCACACG 630
CTAGTGGACC CCGAAGCGGT CTGGGATGCG AACGGGCTG GAACCATAGC GTCAGGGAGG CCGGGGTCCC 700
CCTGGGGCTG CCAGCCCCGG GTGCGAGGAG GCGCGGGGCG AGTGCCAGCC GAAGTCTGCC GTTGCCCAAG 770
AGGCCCCAGG GTGGCGCTGC CCTTGAGCCG GAGCGGACGC CCGTTGGGCA GGGGTCTCGG GCCCACCCGG 840
GCAGGACGCG TGGACCGAGT GACCGTGGTT TCTGTGTGGT GTCACCTGCC AGACCCCGCG AAGAAGCCAC 910
CTCTTTGGAG GGTGCGCTCT CTGGCACGCG CCACTCCAC CCACTCCGTC GCGGCCGCGG CCGCGCGGCG 980
CCCCCATCCA CATCGCGGCC ACCACGTCCT TGGACACGCG CTGTCCCCG GGTGTACGCC GAGACCAAGC 1050
ACTTCTCTTA CTCTCAGGC GACAAGGAGC AGCTGCGGCC CTCTTCTTA CTCAGCTCTC TGAGGCCACG 1120
CCTGACTGGC GCTCGGAGGC TCGTGGAGAC CATCTTCTG GGTTCAGGC CCTGGATGCC AGGGACTCCC 1190
CGCAGGTTCG CCGCCTGCC CCAGCGCTAC TGGCAATGC GCGCCCTGTT TCTGGAGCTG CTTGGGAACC 1260
ACGCGCAGTG CCCCACGGG GTGCTCTCA AGACGCACTG CCGCTGCGA GCTGCGGTCA CCCCAGCAGC 1330
CGGTGTCTGT GCGCGGAGA AGCCCCAGG CTCTGTGGG GCGCCCGAGG AGGAGGACAC AGACCCCGCT 1400
CGCTGTGGTG AGCTGCTCG CCAGCACAGC AGCCCTGGC AGGTGTACGG CTTCGTGCGG GCGTGCCTGC 1470
GCCGCTGGT GCGCCAGGC CTCTGGGCT CCAGGCACAA CGAACGCGCG TTCCTCAGGA ACACCAAGAA 1540
GTTTCATCTC CTGGGAAGC ATGCCAAGCT CTCGCTGACG GAGCTGACGT GGAAGATGAG CGTGGCGGAC 1610
TGCCTTTGGC TGCGCAGGAG CCGAGGGGTT GGCTGTGTT CCGCCGCGA GCACCTCTG CTTGAGGAGA 1680
TCCTGGCCAA GTTCTGTCAC TGGCTGATGA GTGTGTACGT CGTCGAGCTG CTCAGGTCTT TCTTTTATGT 1750
CACGGAGACC ACGTTTCAA AGAACAGGCT CTTTTCTAC CGGAAGAGTG TCTGGAGCAA GTTGCAAGC 1820
ATTGAATCA GACAGCACTT GAAGAGGCT CAGCTCGGG AGCTGTGCGA AGCAGAGTC AGGCAGATC 1890
GGGAAGCCAG GCGCGCCCTG CTGACGTCCA GACTCCGCTT CATCCCCAAG CCTGACGGGC TCGGCGCGAT 1960
TGTGAACATG GACTACGTCG TGGGAGCCAG AACGTTCCG AGAGAAAAGA GGGCCGAGCG TCTCACCTCG 2030
AGGGTGAAG CACTGTTGAG CGTGCTCAAC TACGAGCGG CCGCGCGCCC GCGCCTCTG GCGCCTCTG 2100
TGTGGGCTT GACAGATATC CACAGGGCCT GCGCACCTT CGTGCTGCGT GTGCGGGGCC AGGACCCGCG 2170
GCCTGAGCTG TACTTTGTCA AGGTGGATGT GACGGGCGCG TACGACACCA TCCCCCAGGA CAGGCTCAGC 2240
GAGGTTCATG CCAGCATCAT CAACCCCGAG AACACGTACT CGTGCGCTCG GTATGCGCTG GTCCAGAGG 2310
CCGCCCCATG GCACGTCCCG AAGGCTTCA AGAGCCACGT CTCTACCTTG ACAGACCTCC AGCCGTACAT 2380
GCGACAGTTC GTGGCTCACC TGCAGGAGAC CAGCCCGCTG AGGGATGCGG TCGTCATCGA GCAGAGCTCC 2450
TCCCTGAATG AGGCCAGCAG TGGCCTCTTC GACGCTCTCC TACGCTTCAT GTGCCACCA GCGGTGCGCA 2520
CTAGGGGCAA GTCCACGTC CAGTGCCAGG GGATCCCGCA GGGCTCCATC CTCTCCACG CTCTCTGAG 2590
CCTGTGCTAC GCGCATATG AGAACAGCT GTTTGCGGGG ATTGCGCGGG ACGGGCTGCT CCGCGTTT 2660
GTGGATGATT TCTGTGTTG GACACCTCAC CTCACCCACG CGAAACCTT CCTCAGGACC CTGGTCCGAG 2730
GTGTCCCTGA GTATGGCTGC GTGGTGAAT TCGGAAGAC AGTGGTGAAC TTCCCTGTAG AAGACGAGGC 2800
CCTGGGTGCG ACGGCTTTTG TTCAGATGCC GCGCCACGGC CTATTCCTT GGTGCGGCTT GCTGCTGGAT 2870
ACCCGAGCCC TGGAGGTGCA GAGCGACTAC TCCAGCTATG CCGGACCTC CATCAGAGCC AGTCTCAGCT 2940
TCAACCGCGG CTTCAGGCT GGGAGGAACA TGCCTGCGAA ACTCTTTGG GTTTTGGCGG TGAAGTGTCA 3010
CAGCCTGTTT CTGGATTGCG AGGTGAACAG CCTCCAGACG GTGTGCACCA ACATCTACAA GATCCTCTG 3080
CTGACGGCGT ACAGGTTTCA CGCATGTGTG CTGCACTCC CATTTTCATCA GCAAGTTTGG AAGAACCCCA 3150
GATTTTCTCT GCGCGTCATC TCTGACACGG CCTCCCTCTG CTACTCCATC CTGAAAGCCA AGAACGCGG 3220
GCATTCTCTG TCAAGCTGAC TCGACACCGT GTCACCTACG TGCCACTCCT GGGGTCACTC AGGACAGCCC 3290
AGACCGAGCT GAGTCGGAAG CTCCCGGGGA CGACGCTGAC TGCCCTGGAG GCGCGAGCCA ACCCGGCACT 3360
GCGCTCAGAC TTCAGACCA TCCTGGAGTG ATGGCCACCC GCCCACAGCC AGGCCGAGAG CAGACACCAG 3430
CAGCCCTGTC ACGCCGGGCT CTACGTCCA GGGAGGGAGG GCGCGGCCAC ACCCAGGCC GCACCGCTGG 3500
GAGTCTGAGG CTTGAGTGAG TGTTTGGCG AGGCTGCGAT GTCCGGCTGA AGGCTGAGTG TCCGGCTGAG 3570
GCCTGAGCGA GTGTCCAGCC AAGGGCTGAG TGTCCAGCAC ACCTGCGCTC TTCACTTCCC CACAGGCTGG 3640
CGCTCGGCTC CACCCAGGG CCAGCTTTTC CTCACAGGA GCGCGGCTTC CACTCCCCAC ATAGGAATAG 3710
TCCATCCCCA GATTGCGCAT TGTTCACCCC TCGCCCTGCC CTCCTTTGCC TTCCACCCCC ACCATCCAGG 3780
TGGAGACCCCT GAGAAGGACC CTGGGAGCTC TGGGAATTTG GAGTGACCAA AGGTGTGCCC TGTACACAGG 3850
CGAGGACCCCT GCACCTGGAT GGGGTCTCT GTGGGTCAAA TTGGGGGAG GTGCTGTGGG AGTAAATATC 3920
TGAATATATG AGTTTTCAG TTTTGA AAAA AAAAAA AA 4042

```

Fig. 7

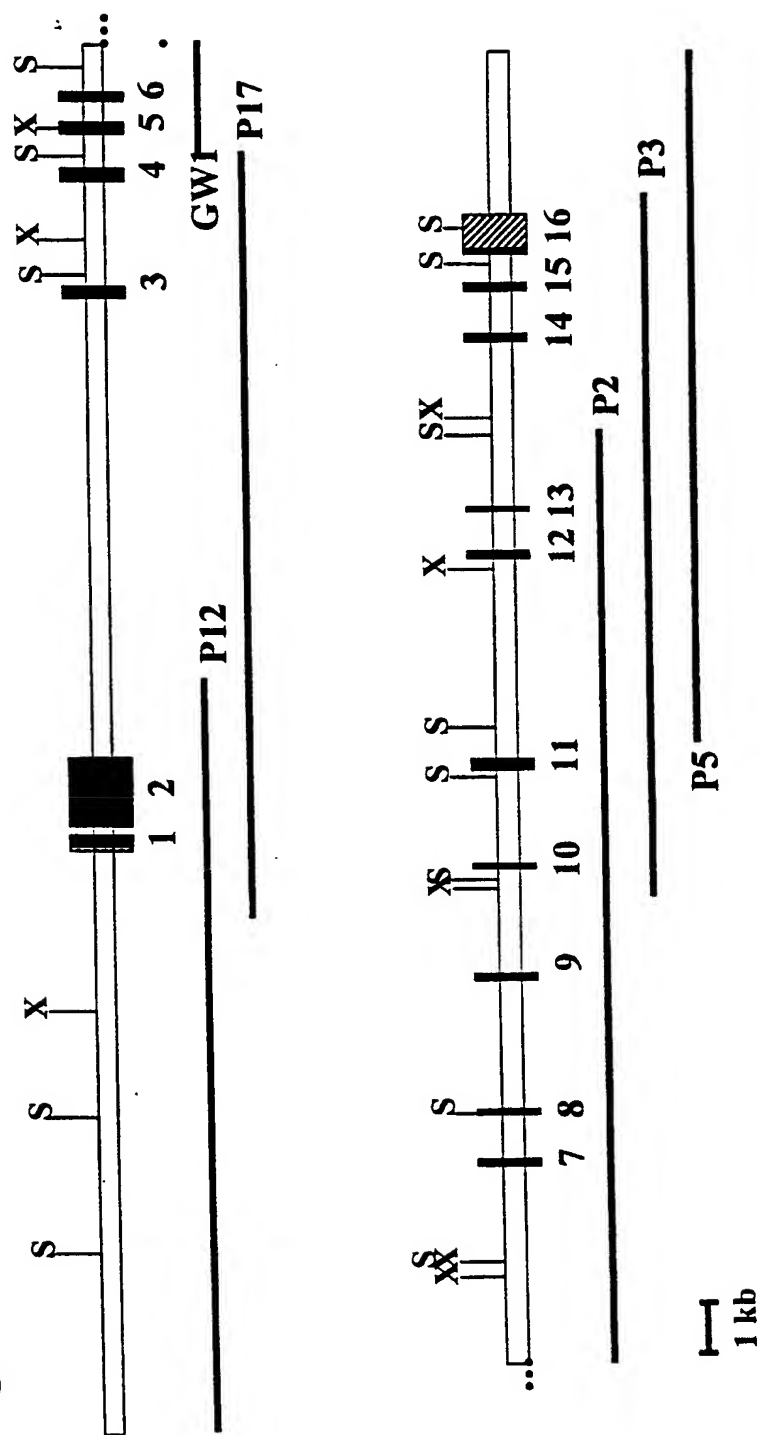


Fig. 8A

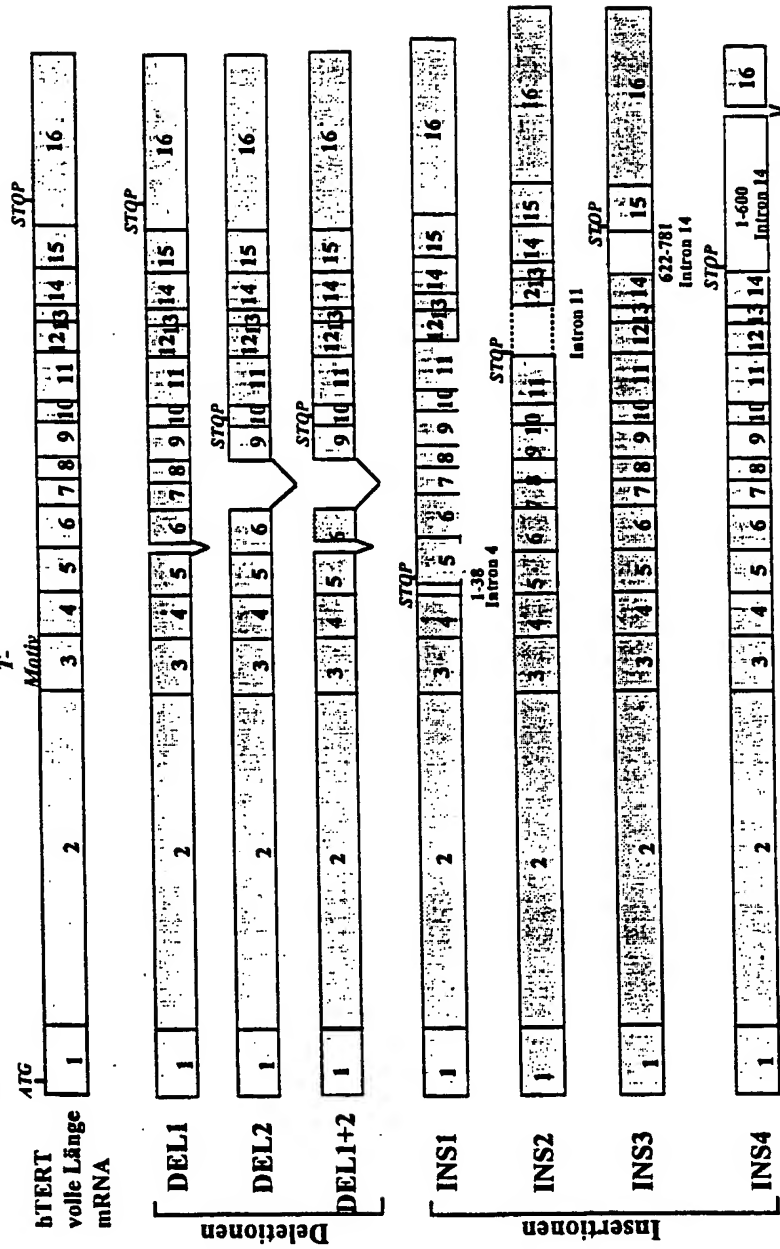


Fig. 8B

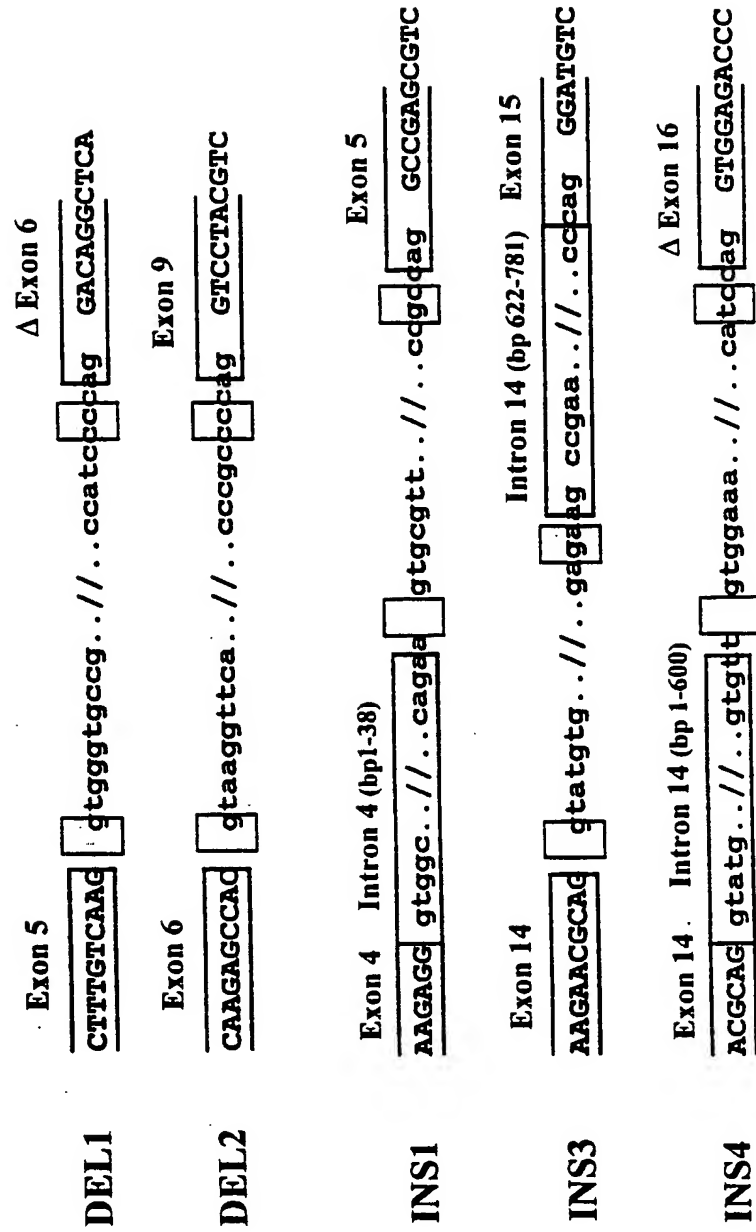


Fig. 9

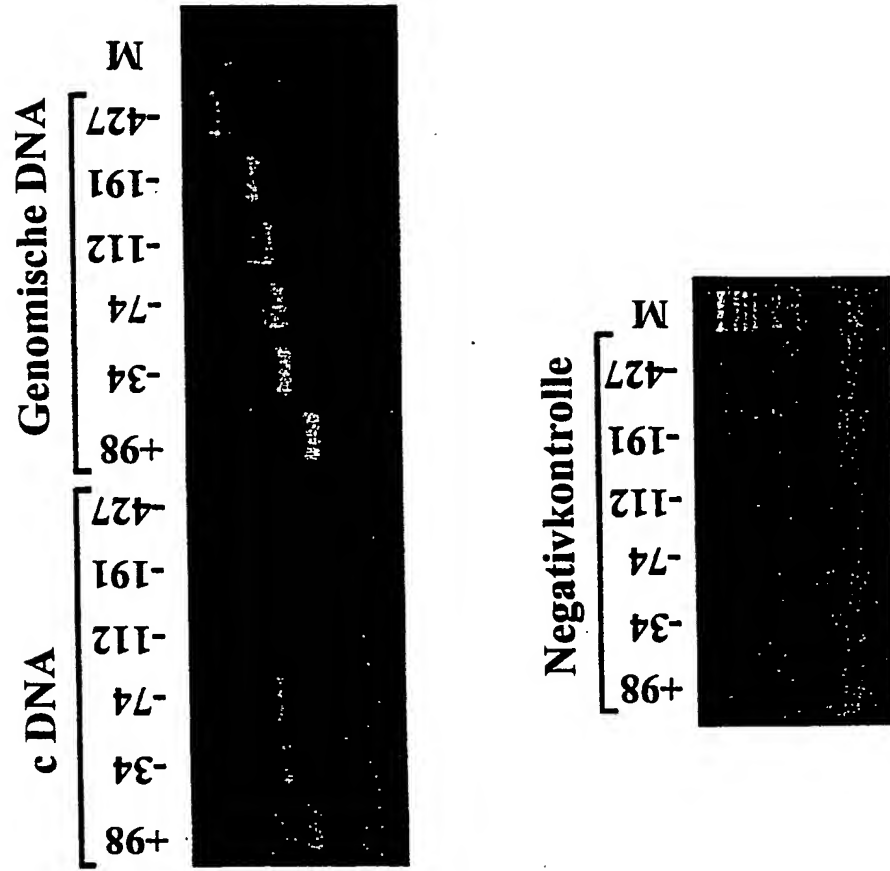


Fig. 10

ACTTGAGCCC AAGAGTTCAA GGCTACGGTG AGCCATGATT GCAACACCAC ACGCCAGCCT TGGTGACAGA -11204
 ATGAGACCCT GTCTCAAAAA AAAAAAAAAA AATTGAAATA ATATAAAGCA TCTTCTCTGG CCACAGTGGA -11134
 ACAAAACCAG AAATCAACAA CAAGAGGAAT TTTGAAACT ATACAAACAC ATGAAATTA AACAAATATAC -11064
 TTCTGAATGA CCAGTGAGTC AATGAAGAAA TTAATAAGGA AATTGAAAAA TTTATTTAAG CAAATGATAA -10994
 CGGAAACATA ACCTCTCAAA ACCCACGGTA TACAGCAAAA GCAGTGCTAA GAAGGAAGTT TATAGCTATA -10924
 AGCAGCTACA TCAAAAAAGT AGAAAAGCCA GGCGCAGTGG CTCATGCCTG TAATCCCGAG ACTTTGGGAG -10854
 GCCAAGGGGG GCAGATCGCC TGAGGTCCAG AGTTCGAGAC CAGCCTGACC AACACAGAGA AACCTTGTCG -10784
 CTACTAAAAA TACAAAATTA GCTGGGCATG GTGGCACATG CCTGTAATCC CAGTACTCG GGAGGCTGAG -10714
 GCAGGATAAC CGCTTGAACC CAGGAGGTGG AGGTTGCGGT GAGCCGGGAT TGCGCCATTG GACTCCAGCC -10644
 TGGGTAACAA GAGTGAACC CTGTCTCAAG AAAAAAAAAA AAGTAGAAAA ACTTAAAAAT ACAACCTAAT -10574
 GATGCACCTT AAAGAAGTAG AAAAGCAAGA GCAAACTAAA CCTAAAAATT GTAAAAAGAA AGAAATTAAT -10504
 AAGATCAGAG CAGAAATAAA TGAAGTGAAG AGATAACAAT ACAAAGATC AACAAAAATTA AAAGTTGGTT -10434
 TTTTGAAGAG ATAAACAAAA TTGACAAACC TTTGCCGAGA CTAAGAAAAA AGGAAAGAA AGCTAAATTA -10364
 ATAAAGTCAG AGATGAAAAA AGAGACATTA CAACTGATAC CACAGAAAT CAAAGGATCA CTAGAGGCTA -10294
 CTATGAGCAA CTGTACACTA ATAAATTGAA AAACCTAGAA AAAATAGATA AATTCTTAGA TGCATACAAC -10224
 CTACCAAGAT TGAACCATGA AGAAATCCAA AGCCCAACCA GACCAATAAC AATAATGGGA TTAAGCCAT -10154
 AATAAAAGT CTCCTAGCAA AGAGAAGCCC AGGACCCAAT GGCTTCCCTG CTGGATTTTA CCAATCATTT -10084
 AAAGAAGAAAT GAATTCGAAT CTTACTCAAA CTATTCTGAA AAATAGAGGA AAGAATACTT CCAAACTCAT -10014
 TCTACATGGC CAGTATTACC CTGATTCCAA AACCAGACAA AAACACATCA AAAACAAACA AACAAAAAA -9944
 CAGAAAGAAA GAAACTACA GGCCAAATATC CCTGATGAAT ACTGATACAA AAATCCTCAA CAAACCACTA -9874
 GCAAAACCAA TTAACAACA CCTTCGAAAG ATCATTCAAT GTGATCAAGT GGGATTTATT CCAAGGATGG -9804
 AAGGATGGTT CAACATATGC AAATCAATCA ATGTGATACA TCATCCCAAC AAAATGAAGT ACAAAACTA -9734
 TATGATTATT TCACTTTATG CAGAAAAAGC ATTTGATAAA ATTTCTGCACC CTTCTAGATA AAAACCCCTCA -9664
 AAAAACCCAG TATACAAGAA ACATACAGGC CAGGCACAGT GGCTCACACC TGCGATCCCA GCCTCTGGG -9594
 AGGCCAAGGT GGGATGATTG CTTGGGCCCC GGAGTTTGAG ACTAGCCTGG GCAACAAAAA GAGACCTGGT -9524
 CTACAAAAAA CTTTTTAA AAATTAGCCA GGCAATGATG CATATGCCTG TAGTCCAGC TAGTCTGGAG -9454
 GCTGAGGTGG GAGATCACT TAAGCCTAGG AGGTCGAGGC TGCACTGAGC CATGAACATG TCACTGTACT -9384
 CCAGCCTAGA CAACGAGACA AGACCCCACT GAATAAGAAG AAGGAGAAGG AGAAGGGAGA AGGGAGGGAG -9314
 AAGGGAGGAG GAGGAGAAGG AGGAGGTGGA GGAGAAGTGG AAGGGGAAGG GGAAGGGAAA GAGGAAGAAG -9244
 AAGAAACATA TTTCAACATA ATAAAAGCCC TATATGACAG ACCGAGGTAG TATTATGAGG TTAAGCCAT -9174
 AGCCTTTCCT CTAAGATCTG GAAAATGACA AGGGCCCACT TTCACCACTG TGATTCAACA TAGTACTAGA -9104
 AGTCCTAGCT AGAGCAATCA GATAAGAGAA AGAAATAAAA GGCATCCAAA CTGGAAAGGA AGAAGTCAA -9034
 TTATCCTGTT TGCAGATGAT ATGATCTTAT ATCTGTAAGA GACTTAAGAC ACCACTAAAA AACTATTAGA -8964
 GCTGAAATTT GGTACAGCAG GATACAAAA CAATGTACAA AAATCAGTAG TATTTCTATA TTCCAACAGC -8894
 AAACAATCTG AAAAAAGAAC CAAAAAGCA GCTACAAATA AAATTAACA GCTAGGAATT AACCAGAGAA -8824
 GTGAAGATG TCTACAATGA AAATAATAA ATGTTGATAA AAGAAATTGA AGAGGGCACA AAAAAAGAA -8754
 AGATATTCCA TGTTCTAGA TTGGAAGAAT AAATCTGTT AAAATGTCCA TACTACCCAA AGCAATTATC -8684
 AAATTCATG CAATCCCTAT TAAATACTA ATGACGTTCT TCACAGAAAT AGAAGAAACA ATTCTAAGAT -8614
 TTGTACAGAA CCACAAAAGA CCCAGAATAG CCAAGCTAT CCTGACCAA AAGAACAAA CTGGAAGCAT -8544
 CACATTACCT GACTTCAAAT TATATACAA AGCTATAGTA ACCCAAACTA CATGGTATG GCATAAAAAAC -8474
 AGATGAGACA TGGACCAGAG GAACAGAATA GAGAATCCAG AAACAAATCC ATGCATCTAC AGTGAACCTA -8404
 TTTTGTACAA AGGTGCCAAG AACATACTTT GGGGAAAAGA TAATCTCTC AATAAATGGT GCTGGAGGAA -8334
 CTGGATATCC ATATGCAAAA TAACAATCT AGAAGCTGT CTCTCACCAT ATACAAAAGC AAATCAAAAT -8264
 GGATGAAAGG CTTAAATCTA AAACCTCAAA CTTTGCAACT ACTAAAAGAA AACACCGGAG AAATCTTCCA -8194
 GGACATTGGA GTGGGCAAG ACTTCTTGAG TAATCCCTG CAGGCACAGG CAACCAAGC AAAAAACAGC -8124
 AAATGGGATC ATATCAAGTT AAAAAGCTTC TGCCAGCAAG AGGAAACAAT CAACAAAGAG AAGAGACAAC -8054
 CCACAGAATG GGAGAATATA TTTGCAAACT ATTCATCTAA CAAGGAATTA ATAACCAGTA TATATAAGGA -7984
 GCTCAAACTA CTCTATAAGA AAAACACCTA ATAAGCTGAT TTTCAAAAT AAGCAAAAGA TCTGGGTAGA -7914
 CATTTCTCAA AATAAGTCAT ACAAATGGCA AACAGGCATC TGAAAAATGT CTCAACACCA CTGATCATCA -7844
 GAGAAATGCA AATCAAACT ACTATGAGAG ATCATCTCAT CCCAGTTAAA ATGGCTTTTA TTCAAAAGAC -7774
 AGGCAATAAC AAATGCCAGT GAGGATGTGG ATAAAAGGAA ACCCTTGGAC ACTGTTGGTG GGAATGGAAG -7704
 TTGCTACCAC TATGGAGAAC AGTTTGAAG TTTCTCAAAA AACTAAAAAT AAAGCTACCA TACAGCAATC -7634
 CCATTGCTAG GTATATACTC CAAAAAGGG AATCAGTGT TCAACAAGCT ATCTCCACTC CCACATTTAC -7564
 TGCAGCACTG TTCATAGCAG CCAAGGTTTG GAAGCAACCT CAGTGTCCAT CAACAGACGA ATGGAAAAAG -7494
 AAAATGTGGT GCACATACAC AATGGAGTAC TACGCAGCCA TAAAAAGAA TGAGATCCTG TCAGTTGCAA -7424
 CAGCATGGGG GGCCTGGTC AGTATGTTAA GTGAAATAAG CCAGGCACAG AAAGACAAAC TTTTCATGTT -7354
 CTCCTTACT TGTGGGAGCA AAAATTAATA CAATTGACAT AGAAATAGAG GAGAATGGTG GTTCTAGAGG -7284
 GGTGGGGGAC AGGGTGACTA GAGTCAACAA TAATTTATTG TATGTTTTAA AATAACTAAA AGAGTATAAT -7214
 TGGGTTGTTT GTAACACAAA GAAAGGATAA ATGCTTGAAG GTGACAGATA CCCATTTTAC CTGATGTGA -7144
 TTATTACACA TTGTATGCCT GTATCAAAAT ATCTCATGTA TGCTATAGAT ATAAACCCTA CTGATTTAAA -7074
 AATTAATAAT TTAATGGCCA GGCACGGTGG CTCATGTCCG TAATCCAGC ACTTTGGGAG GCCGAGGCGG -7004
 GTGGATCACC TGAGGTCAAG AGTTTGAAC CAGTCTGGCC ACCATGATGA AACCTGTCT CTAATAAGA -6934
 TACAAAAATT AGCCAGGCGT GGTGGCACAT ACCTGTAGTC CCAACTACTC AGGAGGCTGA GACAGGAGAA -6864
 TTGCTTGAAC CTGGGAGGCG GAGGTTCGAG TGAGCCGAGA TCATGCCACT GCACGTGACG CTGGGTGACA -6794
 GAGCAAGACT CCATCTCAA ACAAAACAA AAAAAAGAG ATTAAATTTG TAATTTTAT GTACCGTATA -6724
 AATATATACT CTAATATATT AGAAGTTAAA AATTAACA ATTATAAAG GTAATTAACC ACTTAATCTA -6654
 AAATAAGAAC AATGTATGTG GGGTTTCTAG CTTCTGAAGA AGTAAAGTT ATGGCCACGA TGGCAGAAAT -6584

ERSATZBLATT (REGEL 26)

Fig. 10

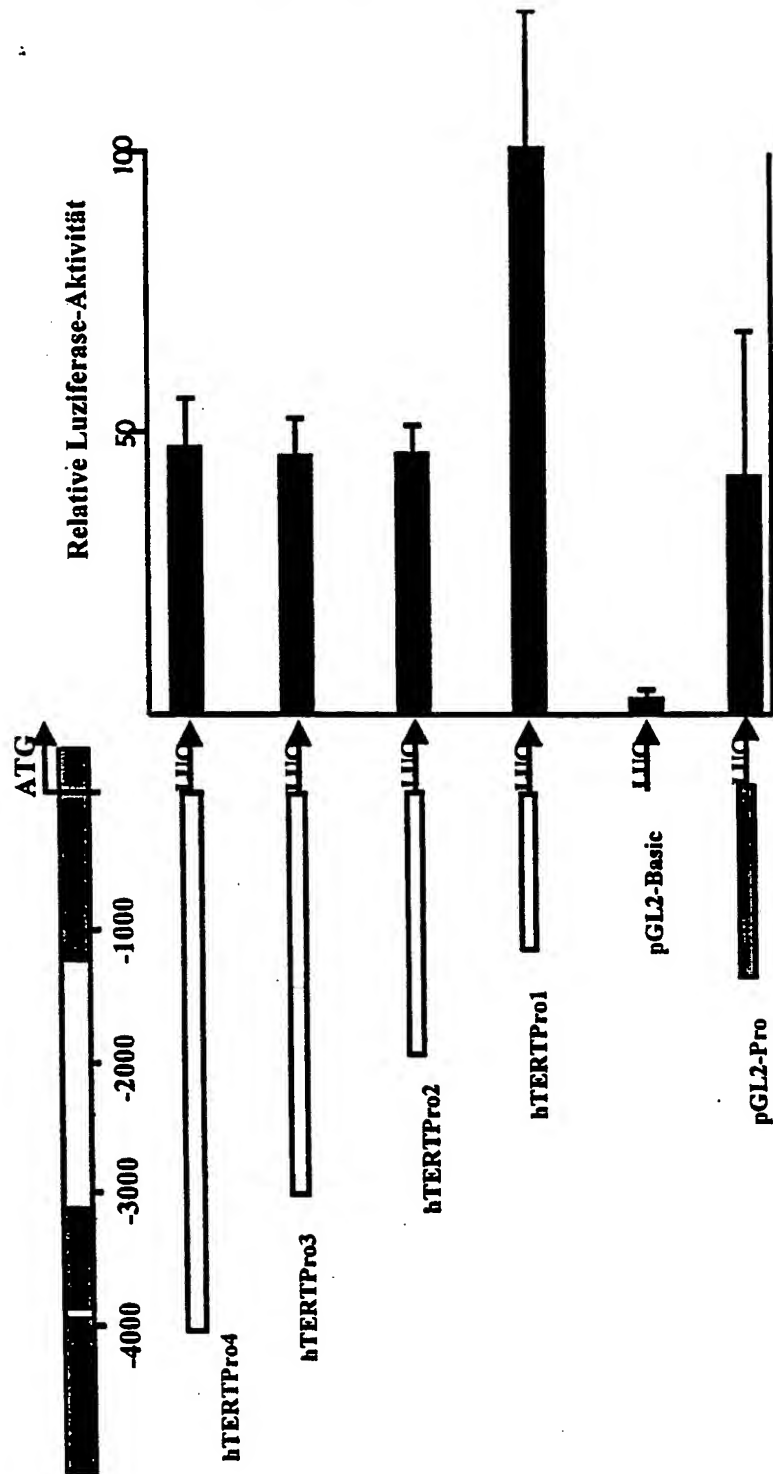
GTGAGGAGGG AACAGTGGAA GTTACTGTGT TTAGACGCTC ATACTCTCTG TAAGTGACTT AATTTTAACC -6514
 AAAGACAGGC TGGGAGAAGT TAAAGAGGCA TTCTATAAGC CCTAAAACAA CTGCTAATAA TGGTGAAAGG -6444
 TAATCTCTAT TAATTACCAA TAATTACAGA TATCTCTAAA ATCGAGCTGC AGAATTGGCA CGTCTGATCA -6374
 CACCGTCCTC TCATTACAGG TGCTTTTTTT CTTGTGTGCT TGGAGATTTT CGATTGTGTG TTCGTGTTTG -6304
 GTTAAACTTA ATCTGTATGA ATCCTGAAAC GAAAAATGGT GGTGATTTCC TCCAGAAAGAA TTAGAGTACC -6234
 TGGCAGGAAG CAGGTGGCTC TGTGGACCTG AGCCACTTCA ATCTTCAAGG GTCTCTGGCC AAGACCCAGG -6164
 TGCAAGGCAG AGGCCTGATG ACCCGAGGAC AGGAAAGCTC GGATGGGAAG GGGCGATGAG AAGCCTGCCT -6094
 CGTTGGTGAG CAGCGCATGA AGTGCCCTTA TTTACGCTTT GCAAAGATTG CTCTGGATAC CATCTGGAAA -6024
 AGGCGGCGAG CGGGAATGCA AGGAGTCAGA AGCCTCCTGC TCAAACCCAG GCCAGCAGCT ATGGCGGCCA -5954
 CCGGGCGCTG TGCCAGAGGG AGAGGAGTCA AGGCACCTCG AAGTATGGCT TAAATCTTTT TTTACCTGA -5884
 AGCAGTGACC AAGGTGTATT CTGAGGGAAG CTTGAGTTAG GTGCCTTCTT TAAACAGAA AGTCATGGAA -5814
 GCACCTTCTT CAAGGGAAAA CCAGACGCCC GCTCTGCGGT CATTACCTC TTTCTCTCTT CCCTCTCTTG -5744
 CCCTCGCGGT TTCTGATCGG GACAGAGTGA CCCCCTGGGA GCTTCTCCGA GCCCGTGTG AGGACCTCT -5674
 TGCAAAAGGGC TCACAGAGCC CCGCCCTGGG AGAGAGGAGT CTGAGCCTGG CTTAATAACA AACTGGGATG -5604
 TGGCTGGGGG CGGACAGCGA CGGCGGGATT CAAAGACTTA ATTCCATGAG TAAATTCAAC CTTTCCACAT -5534
 CCGAATGGAT TTGGAATTTA TCTTAATATT TTCTTAAATT TCATCAAATA ACATTCCAGGA CTGCAGAAAT -5464
 CCAAAGGCGT AAAACAGGAA CTGAGCTATG TTTGCCAAGG TCCAAGGACT TAATAACCAT GTTCAGAGGG -5394
 ATTTTTCGCC CTAAGTACTT TTTATTGGTT TTCATAAGGT GGCTTAGGGT GCAAGGGAAA GTACACGAGG -5324
 AGAGGCCCTG GCGGCAAGGC TATGAGCAGC GCAGGCGCAC CGGGGAGAGA GTCCCCGCC TGGGAGGCTG -5254
 ACAGCAGGAC CACTGACCGT CCTCCCTGGG AGCTGCCACA TTGGGCAACG CGAAGGCGGC CACGCTGCGT -5184
 GTGACTCAGG ACCCCATACC GGCTTCCTGG GCCCACCCAC ACTAACCCAG GAAGTACGCG AGCTCTGAAC -5114
 CCGTGGAAC GAACATGACC CTTGCCTGCC TGCTTCCCTG GGTGGGTCAA GGGTAATGAA TGGGTGTGCA -5044
 GGAATGGCC ATGTAAATTA CACGACTCTG CTGATGGGGA CCGTTCCTTC CATCATTATT CATCTTACC -4974
 CCCAAGGACT GAATGATTCC AGCAACTTCT TCGGGTGTGA CAGGCCATGA CAAAACTCAG TACAAACACC -4904
 ACTCTTTTAC TAGGCCACA GAGCAGGSC CACACCCTG ATATATTAAG AGTCCAGGAG AGATGAGGCT -4834
 GCTTTCAGCC ACCAGGCTGG GGTGACAACA GCGGCTGAAC AGTCTGTTC TCTAGACTAG TAGACCCTGG -4764
 CAGGCACTCC CCCAGATTCT AGGGCTCGGT TGCTGCTTCC CGAGGGCGCC ATCTGCCCTG GAGACTCAGC -4694
 CTGGGGTGCC ACACTGAGGC CAGCCCTGTC TCCACACCCT CCGCTCCAG GCCTCAGCTT CTCCAGCAGC -4624
 TTCTTAAACC CTGGGTGGGC CGTGTCCAG CGCTACTGTC TCACCTGTCC CACTGTGTCT TGTCTCAGCG -4554
 ACGTAGCTCG CACGGTTCCT CCTCACATGG GGTGTCTGTC TCCTTCCCA ACACCTCAT GCGTTGAAGG -4484
 GAGGAGATTC TGCGCCTCCC AGACTGGCTC CTCTGAGCCT GAACCTGGCT CGTGGCCCCC GATGCAGGTT -4414
 CCTGGGCTCC GGCTGCACGC TGACCTCCAT TCCAGGCGC TCCCGTCTC CTGTCATCTG CCGGGGCTG -4344
 CCGGTGTGTT CTTCTGTTTC TGTGCTCCTT TCCACGTCCA GCTGCGTGTG TCTCTGCCG CTAGGGTCTC -4274
 GGGGTTTTTA TAGGCATAGG ACGGGGCGCT GGTGGGCCAG GCGGCTCTTG GGAATGCAA CATTTGGGTG -4204
 TGAAAGTAGG AGTGCCCTGTC CTCACCTAGG TCCACGGGCA CAGGCCTGGG GATGGAGCCC CCGCCAGGGA -4134
 CCGGCCCTTC TCTGCCAGC ACTTTCCTGC CCCCCTCCCT CTGGAACACA GAGTGGCAGT TTCCACAAGC -4064
 ACTAAGCATC CTCTTCCCAA AAGACCCAGC ATTGGCACCC CTGGACATTT GCCCCACAGC CCTGGGAATT -3994

c-Myc
 CACGTGACTA CGCACATCAT GTACACACTC CCGTCCACGA CCGACCCCG CTGTTTTATT TTAATAGCTA -3924
 CAAAGCAGGG AAATCCCTGC TAAATGTCC TTTAACAAAC TGGTTAAACA AACGGGTCCA TCCGCACGGT -3854
 GGACAGTTCC TCACAGTGAA GAGGAACATG CCGTTTATAA AGCCTGCAGG CATCTCAAGG GAATTACGCT -3784
 GAGTCAAAAC TGCCACCTCC ATGGGATACG TACGCAACAT GCTCAAAAG AAGAATTTT ACCCCATGGC -3714
 AGGGGAGTGG TTAGGGGGT TAAGGACGGT GGGGCGGCA GCTGGGGGCT ACTGCACGCA CCTTTTACTA -3644
 AAGCCAGTTT CCTGGTTCTG ATGGTATTGG CTCAGTTATG GGAGACTAAC CATAGGGGAG TGGGGATGGG -3574
 GGAACCCGGA GGCTGTGCCA TCTTTGCCAT GCCCGAGTGT CCTGGGCAGG ATAATGCTCT AGAGATGCC -3504
 ACGTCTGAT TCCCCAAAC CTGTGGACAG AACC CGCCG GCCCCAGGCT CTTTGCAGT GTGATCTCCG -3434
 TGAGGACCCT GAGGTCTGGG ATCCTTCGGG ACTACCTGCA GGGCCGAAA GTAATCCAGG GGTCTGGGA -3364
 AGAGCGGGC AGGAGGGTCA GAGGGGGGCA GCCTCAGGAC GATGGAGGCA GTCAGTCTGA GGCTGAAAAG -3294
 GGAGGGAGGG CCTCGAGCCC AGGCCTGCAA GCGCCTCCAG AAGCTGAAA AAGCGGGGAA GGGACCTCC -3224
 ACGGAGCCTG CAGCAGGAAG GCACGGCTGG CCTTAGCCC ACCAGGGCCC ATCGTGACC TCCGGCTCC -3154
 GTGCCATAGG AGGGCACTCG CGCTGCCCTT CTAGCATGAA GTGTGTGGG ATTTGCAGAA GCAACAGGAA -3084
 ACCCATGCA TGTGAATCTA GGATTATTTC AAAACAAAGG TTTACAGAAA CATCCAAGGA CAGGGCTGAA -3014
 GTGCCCTCCG GCAAGGGCAG GGCAGGCACG AGTGATTTTA TTTAGCTATT TTATTATT TACTTACTTT -2944
 CTGAGACAGA GTTATGCTCT TGTGCCCAG GCTGGAGTGC AGCGGCATGA TCTTGGCTCA CTGCAACCTC -2874
 CGTCTCCTGG GTTCAAGCAA TTCTCGTGCC TCAGCCTCCC AAGTAGCTGG GATTTCAGGC GTGCACCACC -2804
 ACACCCGGCT AATTTTGTAT TTTTAGTAGA GATGGGCTTT CACCATGTTG GTCAAGCTGA TCTCAAAATC -2734
 CTGACCTCAG GTGATCCGCC CACCTCAGCC TCCCAAAGTG CTGGGATTAC AGGCATGAGC CACTGCACCT -2664
 GGCCTATTTA ACCATTTTAA AACTTCCCTG GGCTCAAGTC ACACCCACTG GTAAGGAGTT CATGGAGTTC -2594
 AATTTCCCTT TTAAGCAGGA GTTACCCTCC TTTGATATT TCTGTAATTC TTCTGAGCT GGGGATACAC -2524
 CGTCTCTTGA CATATTCACA GTTCTGTGA CCACCTGTTA TCCCATGGGA CCCACTGCAG GGGCAGCTGG -2454
 GAGGCTGCAG GCTTCAGGTC CCAGTGGGTT TGCCATCTGC CAGTAGAAAC CTGATGTAGA ATCAGGGCGC -2384
 AAGTGTGGAC ACTGTCTGA ATCTCAATG CTCAGTGTGT GCTGAAACAT GTAGAAATTA AAGTCCATCC -2314
 CTCCTACTCT ACTGGGATTG AGCCCCCTCC CTATCCCCC CCAGGGGCG AGGAGTTCCT CTCACTCTG -2244
 TGGAGGAAG AATGATACTT TGTATTTTT CACTGCTGCT ACTGAATCCA CTGTTTCATT TGTGTTTG -2174
 TTTGTTTGT TTTGAGAGC GGTTCCTCT TGTGTGCTGA GGCTGGAGG AGTGCAATGG CGCGATCTTG -2104
 GCTTACTGCA GCCTCTGCCT CCCAGGTTCA AGTGATTCTC CTGCTTCCG CTCCCATTTG GCTGGGATTA -2034
 CAGGCACCCG CCACCATGCC CAGCTAATTT TTTGTATTT TAGTAGAGAC GGGGGTGGG GGGGTTACC -1964

Fig. 10

ATGTTGGCCA GGCTGGTCTC GAACTTCTGA CCTCAGATGA TCCACCTGCC TCTGCCTCCT AAAGTGCTGG -1894
 GATTACAGGT GTGAGCCACC ATGCCCAGCT CAGAATTTAC TCTGTTTGA AACATCTGGG TCTGAGGTAG -1824
 GAAGCTCACC CCACTCAAGT GTTGTGGTGT TTTACAAT-BoxGCCAATGATAGAATT TTTTATTGT TGTTAGAACA -1754
 CTCTTGATGT TTTACACTGT GATGACTAAG ACATCATCAG CTTTTCAAAG ACACACTAAC TGCACCCATA -1684
 ATACTGGGGT GTCTTCTGGG TATCAGCAAT CTTCATTGAA TGCCGGGAGG CGTTTCCTCG CCATGCACAT -1614
 GGTGTTAATT ACTCCAGCAT AATCTTCTGC TTCCATTCT TCTCTTCCCT CTTTAAAAAT TGTGTTTTCT -1544
 ATGTTGGCTT CTCTGCAGAG AACCAGTGTA AGCTACAAC TAACTTTGT TGGAACAAAT TTTCCAACC -1474
Spl
GCCCTTTGC CCTAGTGGCA GAGACAATTC ACAAACACAG CCCTTTAAAA AGGCTTAGGG ATCACTAAGG -1404
 GGATTTCTAG AAGAGCGACC TGTAATCCTA AGTATTTACA AGACGAGGCT AACCTCCAGC GAGCGTGACA -1334
 GCCCAGGGAG GGTGCGAGGC CTGTTCAAAT GCTAGCTCCA TAAATAAAGC AATTTCTCTC GGCAGTTTCT -1264
 GAAAGTAGGA AAGGTTACAT TTAAGTTGC GTTGTGTAGC ATTCAGTGT TTGCCGACCT CAGCTACAGC -1194
 ATCCCTGCAA GGCCTCGGGA GACCCAGAAG TTTCTCGCCC CTTAGATCC AACTTGAGC AACCCGGAGT -1124
 CTGGATTCTT GGGAAAGTCTT CAGCTGTCCT GCGGTTGTGC CGGGGCCCCA GGTCTGGAGG GGACCAAGTG -1054
 CCGTGTGGCT TCTACTGCTG GGCTGGAAGT CGGGCTCCT AGCTCTGCAG TCCGAGGCTT GGAGCCAGGT -984
 GCCTGGACCC CGAGGCTGCC CTCCACCTG TGCGGGCGGG ATGTGACCAG ATGTTGGCCT CATCTGCCAG -914
 ACAGAGTGCC GGGGCCCAGG GTCAAGGCCG TTGTGGCTGG TGTGAGGCGC CCGGTGCGCG GCCAGCAGGA -844
 GCGCCTGGCT CCATTCCAC-BoxCCCACCCTTCTCG ACGGGASplCGCCCGTTGGT GATTACAGA TTTGGGTGG -774
 TTTGCTCATG GTGGGACCC CTCGCCGCT GAGAACCTGC AAAGAGAAAT GACGGGCTG TGTCAAGGAG -704
 CCCAAGTCGC GGGGAAGTGT TGCAGGAGG CACTCCGGA GGTCCCGCT GCCCGTCCAG GGAGCAATGC -634
 GTCCTCGGGT TCAP-2TCCCCAGCCGCTCTAC GCGCCTCCGT CCTCCCTTC ACGTCCGGCA TTCGTGGTGC -564
 CCGGAGCCCG ACGCCCCGCG TCCGGACCTG GAGGCAGCCC TGGGTCTCCG GATCAGGCCA GCGGCCAAAG -494
 GGTGCGCGCA CGCACCTGTT CCCAGGGCCT CCACATCATG GCCCTCCCT CGGGTTACCC CACAGCCTAG -424
 GCCGATTGCA CCTCTCTCCG CTGGGGCCCT CGCTGGCGTC CCTGCACCCT GGGAGCGCGA GCGGSplGCGCG -354
GCGGGGAAG CGCGGCCAG ACCCCCGGGT SplCCGCCGGAG CAGCTGCGCT GTCGGGGCCA GGCCGGGCTC -284
 CCAGTGGATT CGCGGGCACA GACGCCCAGG ACCGCGCTCC c-MycCCACGTGCG GAGGGACTGG GGACCCGGGC -214
 ACCCGTCTG CCCCTTACC TTCCAGCTCC GCCTCTCCG CGCGGACSplCCGCCCGTCCC GACCCCTCCC -144
 GGGTCCCCGG CCCAGCCCCC TCCGGGCCCT CCCAGCCCT CCCCTTCTT TCCGCGGCCSplCGCCCTCTCC -74
 TCGCGCGCG AGTTTACAGC AGCGCTGCG c-MycACGTGGGAAG CCCTGGCCCC GGCCACCCCC -4
 GCGATG

Fig.: 11



SEQUENZPROTOKOLL

<110> Bayer AG

5 <120> Regulatorische DNA-Sequenzen aus der 5i-Region vom Gen
der humanen katalytischen Telomerase-Untereinheit und
deren diagnostische und therapeutische Verwendung

10 <130> LeA32805-Ausland

<140>

<141>

<160> 20

15 <170> PatentIn Vers. 2.0

<210> 1

<211> 5126

20 <212> DNA

<213> Homo sapiens

<400> 1

25 gagctctgaa ccgtggaaac gaacatgacc cttgcctgcc tgcttccctg ggtgggtcaa 60
gggtaataaa gtggtgtgca ggaaatggcc atgtaaaatta cagcactctg ctgatgggga 120
ccgttccttc catcattatt catcttcacc cccaaggact gaatgattcc agcaacttct 180
tcgggtgtga caagccatga caaaactcag tacaacacc actcttttac taggcccaca 240
gagcacgggc cacacccctg atataatga agtccaggag agatgaggct gctttcagcc 300
accaggctgg ggtgacaaca gcggctgaac agtctgttcc tctagactag tagaccctgg 360
30 caggcactcc cccaaattct agggcctggt tgcctgttcc cgaggggccc atctgcctcg 420
gagactcagc ctgggggtgcc acactgaggc cagccctgtc tccacacccct ccgcctccag 480
gcctcagctt ctccagcagc ttccctaaacc ctgggtgggc cgtgttccag cgctactgtc 540
tcacctgtcc cactgtgtct tgcctcagcg acgtagctcg caggggttct cctcacatgg 600
ggtgtctgtc tcttcccca acactcacat gcgttgaagg gaggagattc tgcgcctccc 660
35 agactggctc ctctgagcct gaacctggct cgtggccccc gatgcagggt cctggcgctc 720
ggctgcagc tgacctccat ttccaggcgc tcccgtctc ctgtcatctg ccggggcctg 780
ccggtgtgtt cttctgtttc tgcctcctt tccacgtcca gctgcgtgtg tctctgccc 840
ctagggcttc ggggttttta taggcataag acgggggcgt ggtggggcag ggcgctctt 900
ggaaatgcaa catttgggtg tgaagttagg agtgcctgtc ctacactagg tccacgggca 960
40 caggcctggg gatggagccc ccgccaggga ccgcccttc tctgcccagc actttcctgc 1020
ccccctccct ctggaacaca gattggcagt ttccacaagc actaagcacc ctctcccaa 1080
aagaccagc attggcacc ctggacattt gccccacagc cctgggaatt cactgtacta 1140
cgcacatcat gtacacactc ccgtccacga ccgacccccg ctgttttatt ttaatagcta 1200
caaagcagg aaatccctgc taaaatgtcc ttaacaaac tgggttaaca aacgggtcca 1260
45 tccgcacggt ggacagtcc tcacagtga gaggaacatg ccgtttataa agcctgcagg 1320
catctcaagg gaattacgct gagtcaaac tgccacctcc atgggatacg tacgcaacat 1380
gctcaaaaag aaagaatttc accccatggc aggggagtg ttaggggggt taaggacggt 1440
ggggggcgca gctgggggct actgcacgca ccttttacta aagccagttt cctggttctg 1500
atgggtattg ctacgttatg ggagactaac cataggggag tggggatggg ggaacccgga 1560
50 ggctgtgcca tctttgcat gcccgagtgt cctgggcagg ataatgctc agagatgcc 1620
acgtcctgat tccccaaa ctgtggacag aacccgcccg gcccagggc ctttgagggt 1680
gtgatctccg tgaggacct gaggtctggg atccttcggg actacctgca ggcccgaaaa 1740
gtaatccagg ggttctggga agaggcgggc agggagggtc agggggggca gcctcaggac 1800
gatggaggca gtcagtctga ggctgaaaag ggaggaggg cctcgagccc aggcctgcaa 1860
55 gcgcctccag aagctgaaa aagcggggaa gggaccctcc acggagcctg cagcaggaa 1920
gcacggctgg cccttagccc accaggggcc atcgtggacc tccggcctcc gtgccatagg 1980
agggcactcg cgctgccctt ctacatgaa gtgtgtggg atttgagaa gcaacaggaa 2040
acccatgcac tgtgaatcta ggattatttc aaacaaagg ttacagaaa catccaagga 2100
cagggtgaa gtgcctccg gcaaggcgag ggcaggcac agtgatttta tttagctatt 2160
60 ttattttatt tacttacttt ctgagacaga gttatgctt tgttggccag gctggagtgc 2220
agcggcatga tcttggctca ctgcaacctc cgtctcctgg gttcaagcaa ttctcggtcc 2280
tcagcctccc aagtagctgg gatttcaggc gtgcaaccac acaccggct aattttgtat 2340
tttagtaga gatgggcttt caccatgtt gtcaagctga tcccaaaatc ctgacctcag 2400
gtgatccgc caccctcagc tccaaaagt ctgggattac aggcattgag cactgcacct 2460

```

ggcctattta accattttta aacttccctg ggctcaagtc acaccactg gtaaggagtt 2520
catggagttc aattttccct ttactcagga gttaccctcc ttgatattt tctgtaattc 2580
ttcgtagact ggggatacac cgtctcttga catattcaca gtttctgtga ccacctgtta 2640
tcccatggga ccactgcag gggcagctgg gaggtgcag gcttcaggtc ccagtggggt 2700
5 tgcctatctgc cagtagaaac ctgatgtaga atcagggcgc aagtgtggac actgtcctga 2760
atctcaatgt ctcatgtgtg gctgaaacat gtagaaatta aagtcacatc ctctactctc 2820
actgggattg agccccctcc ctatccccc ccaggggcag aggagttcct ctactcctg 2880
tggagggaag aatgatactt tgttattttt cactgctggg actgaatcca ctgtttcatt 2940
tgttggtttg tttgttttgt tttagagagg ggtttcactc ttgttgctca ggctggaggg 3000
10 agtgcaatgg cgcgatcttg gcttactgca gcctctgcct ccaggttca agtgattctc 3060
ctgcttccgc ctcccatttg gctgggatta caggcacccg ccacatgcc cagctaattt 3120
tttgtatttt tagtagagac ggggggtggg ggggttcacc atgttgcca ggctgggtctc 3180
gaacttttga cctcagatga tccacctgcc tctgcctcct aaagtgcctg gattacaggt 3240
gtgagccacc atgcccagct cagaatttac tctgtttaga aacatctggg tctgaggtag 3300
15 gaagctcacc ccactcaagt gttgtgggtg tttaagccaa tgatagaatt tttttattgt 3360
tgtagaaca ctcttgatgt ttacactgt gatgactaag acatcatcag cttttcaaag 3420
acacactaac tgcaccata atactggggg gtcttctggg tatcagcaat cttcattgaa 3480
tgcggggagg cgtttcctcg ccctgcacat ggtgttaatt actccagcat aatcttctgc 3540
ttccatttct tctcttccct cttttaaaat tgtgttttct atgttggtct ctctgcagag 3600
20 aaccagtgtg agtacaact taacttttgt tggaaacaaat ttccaaaacc gccctttgc 3660
cctagtggca gagacaattc acaaacacag ccttttaaaa aggcttaggg atcactaagg 3720
ggatttctag aagagcgacc tgtaatccta agtatttaca agacagggtc aacctccagc 3780
gagcgtgaca gccaggggag ggtgcgaggc ctgttcaaat gctagctcca taaataaagc 3840
aatttccctc ggaggtttct gaaagtagga aagggttacct ttaagggtgc gttttgttagc 3900
25 atttcagtgt ttgcccagct cagctacagc atccctgcaa ggcctcggga gaccagaag 3960
tttctcgcct ccttagatcc aaacttgagc aacccggagt ctggattcct gggaagcct 4020
cagctgtcct gcggttgtgc cggggcccca ggtctggagg ggaccagtgg ccgtgtggct 4080
tctactgtg ggctggaagt cgggcctcct agctctgcag tccgaggett ggagccaggt 4140
gcttggaacc cgaggtgccc ctccacctg tgcgggcggg atgtgaccag atgttgacct 4200
30 catctgccag acagagtgcg gggggccagg gtcaaggccg ttgtggctgg tgtgaggcgc 4260
ccggtgcgcg gccagcaggg gcgcctggct ccatttccca cctttctctg accggaccgc 4320
cccgggtggg gattaacaga tttgggtggg ttgtctcatg gtgggaacc ctgcgcgcct 4380
gagaacctgc aaagagaaat gacgggcctg tgtcaaggag cccaagtgcg ggggaagtgt 4440
tgacgggagg cactccggga ggtcccgctg gccctccag ggaagcaatgc gtctcgggt 4500
35 tcgtcccccag ccggtctctc gcgcctcctg cctcccttc acgtccggca ttctgtgtgc 4560
ccggagcccg acgccccgcg tccggacctg gaggcagccc tgggtctccg gatcaggcca 4620
gcggccaaaag ggtcgcgcga cgcacctgtt cccagggcct ccacatcatg gccctcctc 4680
cgggttacc cccagcctag gccgattcga cctctctccg ctggggccct cgttggcgct 4740
cctgcaccct gggagcgcgga gcggcgcgcg ggcggggaag cgcggcccag acccccgggt 4800
40 ccgcccggag cagctgcgct gtcggggcca ggccgggctc ccagtggatt cgcgggcaca 4860
gacgcccagg accgcgctcc ccacgtggcg gagggactgg ggaccgggc acccgtcctg 4920
ccccttcacc ttccagctcc gcctcctccg cgcggacccc gccctgtccc gacccctccc 4980
gggtcccccg ccagccccc tccgggcccct cccagccccct ccccttctct tccgcggccc 5040
gcccctctcc tccggcgcg agtttccagg agcgtctgct cctgctgcgc acgtgggaag 5100
45 ccttgcccc ggccaccccc gcgatg 5126

```

<210> 2

<211> 4042

<212> DNA

50 <213> Homo sapiens

<400> 2

```

gtttcaggca gcgctgcgct ctgctgcgca cgtgggaagc cctggccccg gccacccccg 60
cgatgccgcg cgtccccgcg tgcgagccg tgcgctccct gctgcgcagc cactaccgcg 120
55 aggtgctgcc gctggccacg ttctgtgcgc gcctggggcc ccagggtctg cggctggtgc 180
agcgcgggga cccggcggtt ttccgcgcgc tgggtggcca gtgcctggtg tgcgtgcctc 240
gggacgcagc gccgcccccc gccgccccct ccttccgcca ggtgtcctgc ctgaaggagc 300
tgggtggccc agtgcctgag aggtctgtgc agcgcggcgc gaagaactgt ctggccttcg 360
60 gcttcgcgct gctggacggg gcccgcgggg gcccccccg ggccttcacc accagcgtgc 420
gcagctacct gcccacacg gtgaccgacg cactgcgggg gagcggggcg tgggggctgc 480
tgctgcgcgc cgtgggcgac gacgtgctgg ttacactgct ggcacgctgc gcgtctcttg 540
tgctggtggc tcccagctgc gcctaccagg tgtgcgggcc gccgtgtac cagctcggcg 600
ctgccactca gggccggccc ccgcccacag ctagtggacc ccgaaggcgt ctgggatgcg 660
aaggggcctg gaaccatagc gtcagggagg ccggggctcc cctgggacct ccagccccgg 720
65 gtgcgaggag gcgcgggggc agtgcagcc gaagtctgct gttgcccag aggccacggc 780

```


3 / 18

gtggcgctgc ccctgagccg gagcggacgc ccgttgggca ggggtcctgg gcccaccggy 840
 gcaggacgcy tggaccgagt gaccgtgggt tctgtgtggt gtcacctgcc agaccgcgcg 900
 aagaagccac ctcttggag ggtgcgctct ctggcacgcy ccactccac ccatccgtgy 960
 gccgccagca ccacgcgggc ccccatcca catcgcgcc accacgtccc tgggacacgc 1020
 5 cttgtccccc ggtgtacgcc gagaccaagc acttcctcta ctctcaggc gacaaggagc 1080
 agctgcggcc ctccctccta ctcagctctc tgaggccag cctgactggc gctcgaggc 1140
 tctgtggagac catctttctg ggttccaggc cctggatgcc agggactccc cgcaggttgc 1200
 cccgcctgcc ccagcgctac tggcaaatgc ggccctgtt tctggagctg cttgggaacc 1260
 acgcgcagtg cccctacggg gtgctcctca agacgcactg cccgctgcga gctgcggtca 1320
 10 cccacgacgc cgtgtgtctg gcccgggaga agccccagg cctgtgtggc gccccgagg 1380
 aggaggacac agacccccgt cgcctgggtg agctgctccg ccagcacagc agccccggc 1440
 aggtgtacgg ctctgtgcgg gctgcctgc gccggtggt gcccccaggc ctctggggct 1500
 ccaggcacaa cgaacgcgc ttcctcagga acaccaagaa gttcatctcc ctggggaagc 1560
 atgccaaagt ctgcgtcgag gagctgacgt ggaagatgag cgtgcgggac tgcgcttggc 1620
 15 tgcgcaggag cccagggtt ggtgtgttc cggccgcaga gcaccgtctg cgtgaggaga 1680
 tcttgcccaa gttcctgcac tggctgatga gtgtgactg cgtcgagctg ctccaggtctt 1740
 tcttttatgt caccggagac acgtttcaaa agaacaggct cttttctac cgggaagagt 1800
 tctggagcaa gttgcaaaagc attggaatca gacagcactt gaagagggtg cagctgcggg 1860
 agctgtcgga agcagaggtc aggcagcatc gggaagccag gccgcctctg ctgactgcca 1920
 20 gactccgctt catccccaag cctgacgggc tgcggccgat tgtgaacatg gactacgtcg 1980
 tgggagccag aacgttccgc agagaaaaga gggccgagcg tctcacctcg agggggaagg 2040
 cactgttcag cgtgctcaac tacgagcggg cgcggcgccc cggcctcctg ggcgcctctg 2100
 tgctgggcct ggacgatata cacagggcct ggccgacctt cgtgctgctg gtgcgggccc 2160
 aggaaccgcc gcctgagctg tactttgtca aggtggatgt gacgggcgcg tacgacacca 2220
 25 tccccagga caggctcacg gaggtcatcg ccagcatcat caaacccag aacacgtact 2280
 gcgtgctcg gtatgccgtg gtccagaagg ccgcccatgg gcacgtccgc aaggccttca 2340
 agagccacgt ctctacctt acagacctcc agccgtacat gcgacagttc gtggctcacc 2400
 tgcaaggagc cagccccgtg agggatgccc tgcgtatcga gcagagctcc tccctgaatg 2460
 aggccagcag tggcctcttc gacgtcttcc tacgcttcat gtgccaccac gccgtgcga 2520
 30 tcaggggcaa gtccctacgtc cagtgccagg ggatcccga gggctccatc ctctccacgc 2580
 tgctctcgag cctgtgctac ggcgacatgg agaacaagct gtttgcgggg attcggcggg 2640
 acgggctgct cctgcgtttg gtggatgatt tcttgttggg gacacctcac ctacccacg 2700
 cgaaaacctt cctcaggacc ctggctcgag gtgtccctga gtatggctgc gtggtgaact 2760
 tgcggaagac agtgggtaac ttcctctgag aagacgaggc cctgggtggc acggctttt 2820
 35 ttcagatgcc ggcacacggc ctattcccc ggtgcggcct gctgctggat acccggaacc 2880
 tggaggtgca gagcgactac tccagctatg cccggacctc catcagagcc agtctcact 2940
 tcaaccgcgg ctccaaggct gggaggaaca tgcgtcgcaa actctttggg gtcttgcggc 3000
 tgaagtgtca cagctgttt ctggatttgc aggtgaacag cctccagacg gtgtgacca 3060
 40 acatctacaa gatectctg ctgcaggcgt acaggtttca cgcattgtgt ctgcagctcc 3120
 catttcatca gcaagtgttg aagaacccca catttttctt gcgcgtcatc tctgacacgg 3180
 cctccctctg ctactccatc ctgaaagcca agaagcagg gatgtcgctg ggggccaagg 3240
 gcgcgcggcg cctctgccc tccgaggcgg tgcatggctg gtgccacca gcatctctgc 3300
 tcaagctgac tcgacaccgt gtcacctacg tgccactcct ggggtcactc aggcagaccc 3360
 agacgcagct gagtgcgaag ctcccgggga gcagctgac tgccctggag gccgcagcca 3420
 45 acccggcact gccctcagac ttcaagacca tccctggactg atggccaccc gccacagcc 3480
 agggcgagag cagacaccag cagccctgtc acgcccggct ctacgtccca gggaggaggg 3540
 ggccgcccac acccaggccc gcaccgctgg gagtctgagg cctgagttag tgtttggccc 3600
 aggcctgcat gtccggctga aggtgagtg tccggctgag gcctgagcga gtgtccagcc 3660
 50 aagggtgag tgtccagcac acctgcccgtc ttcaacttccc cacaggctgg cgtcgggtc 3720
 caccacaggg ccagcttttc ctaccagga gcccggttc cactccccac ataggaaatg 3780
 tccatcccca gattcgccat tgttcacccc tcgcccgtgc ctcttttggc ttccaccccc 3840
 accatccagg tggagaccct gagaaggacc ctgggagctc tgggaatttg gagtgaacaa 3900
 aggtgtgccc tgtacacagg caggagacct gcacctggat ggggtccct gtgggtcaaa 3960
 55 ttggggggag gtgctgtggg agtaaaatag tgaatatatg agtttttcag ttttgaaaaa 4020
 aaaaaaaaaa aaaaaaaaaa aa 4042

<210> 3
 <211> 11276
 <212> DNA
 60 <213> Homo sapiens

<400> 3
 acttgagccc aagagttcaa ggctacgggtg agccatgatt gcaacaccac acgcagcct 60
 65 tgggtgacaga atgagaccct gtctcaaaaa aaaaaaaaaa aattgaaata atataaagca 120
 tcttctcttg ccacagtgga acaaaaccag aaatcaacaa caagaggaat tttgaaact 180

atacaaacac atgaaaatta aacaatatac ttctgaatga ccagtgaagtc aatgaagaaa 240
 ttaaaaaggga aattgaaaaa ttattttaag caaatgataa cggaacata acctctcaaa 300
 acccacggta tacagcaaaa gcagtgtctaa gaagggaagt tatagctata agcagctaca 360
 tcaaaaaagt agaaaagcca ggcgcagtggt ctcctgcctg taatcccagc actttgggag 420
 5 gccaaaggcg gcagatcgcc tgaggtcagg agttcgagac cagcctgacc aacacagaga 480
 aacctgtctg ctactaaaaa tacaaaatta gctgggcatg gtggcacatg cctgtaatcc 540
 cagctactcg ggaggctgag gcaggataac cgcttgaacc caggagggtg aggttgcggt 600
 gagccgggat tgcgccattg gactccagcc tgggtaacaa gagtgaacc ctgtctcaag 660
 aaaaaaaaaa aagtagaaaa acttaaaat acaacctaat gatgcacct aaagaactag 720
 10 aaaaagcaaga gcaaaactaaa cctaaaattg gtaaaagaaa agaaataata aagatcagag 780
 cagaaataaa tgaactgaa agataacaat acaaaagatc aacaaaatta aaagtgtggt 840
 ttttgaaaag ataaacaaaa ttgacaaacc tttgccaga ctaagaaaaa aggaaagag 900
 acctaaataa ataaagtcag agatgaaaaa agagacatta caactgatac cacagaaatt 960
 caaaggatca ctagggtcta ctatgagcaa ctgtacacta ataaattgaa aaacctagaa 1020
 15 aaaatagata aattcctaga tgcatacaac ctaccaagat tgaacctga agaaatccaa 1080
 agcccaaca gaccaataac aataatggga ttaaagccat aataaaaagt ctcttagcaa 1140
 agagaagccc aggaccaat ggcttccctg ctggatttta ccaatcattt aaagaagaa 1200
 gaattccaat cctactcaaa ctattctgaa aaatagagga aagaatactt ccaaactcat 1260
 tctacatggc cagtattacc ctgattccaa aaccagacaa aaacacatca aaaaacaaaca 1320
 20 aacaaaaaaa cagaaagaaa gaaaactaca ggccaatatc cctgatgaat actgatacaa 1380
 aaatcctcaa aaaaacataa gcaaaccaaa ttaaacacaa ccttcgaaag atcatctatt 1440
 gtgarcaagt gggatttatt ccagggtatg aaggatggtt caacatattc aaatcaatca 1500
 atgtgataca tcatcccaac aaaatgaagt acaaaaacta tatgattatt tcaactttatg 1560
 cagaaaaagc atttgataaa attctgcacc ctctatgata aaaaacctca aaaaaccagg 1620
 25 tatacaagaa acatacaggg caggcacagt ggctcacacc tgcgatccca gcaactctggg 1680
 aggccaaagt gggatgatg ctggggccca ggagtttgag actagccttg gcaacaaaat 1740
 gagacctggc ctacaaaaaa cttttttaa aaatagcca ggcattgatg catatgcctg 1800
 tagtcccagc tagtctggag gctgaggtgg gagaatcact taagcctagg aggtcgaggc 1860
 tgcagtgaac catgaacatg tcaactgtact ccagcctaga caacagaaca agacccact 1920
 30 gaataaagag aaggagaagg agaagggaaga agggagggag aaggagaggag gaggagaagg 1980
 agggaggtgga ggagaagtgg aaggggaagg ggaagggaag gaggaagag aagaacata 2040
 tttcaacata ataaaagccc tatatgacag accgaggtag tattatgagg aaaaactgaa 2100
 agcctttcct ctaagatctg gaaaatgaca agggccact ttcaccactg tgattcaaca 2160
 tagtactaga agtctagct agagcaatca gataagagaa agaaataaaa ggcattccaa 2220
 35 ctggaaagga agaagtcaaa ttatcctgtt tgcagatgat atgactttat acttggaaga 2280
 gacttaagac accactaaaa aactattaga gctgaaattt ggtacagcag gatacaaaat 2340
 caatgtacaa aaatcagtag tatttctata ttccaacagc aaacaatctg aaaaagaaac 2400
 caaaaaagca gctacaaata aaattraaaa gctaggaatt aaccaaagaa gtgaaagatc 2460
 tctacaatga aaactataaa atgttgataa aagaaattga agagggcaca aaaaagaaa 2520
 40 agatattcca tgttcataga ttggaagaat aaatactgtt aaaatgtcca tactacccaa 2580
 agcaatttac aaattcaatg caatccctat taaaatacta atgacgttct tcacagaaat 2640
 agaagaaaca attctaagat ttgtacagaa ccacaaaaga cccagaatag ccaaagctat 2700
 cctgaccaa aagaacaaaa ctggaagcat cacattacct gacttcaaat tatactacaa 2760
 agctatagta acccaacta catggtactg gcataaaaac agatgagaca tggaccagag 2820
 45 gaacagaata gagaatccag aaacaaatcc atgcatctac agtgaactca tttttgacaa 2880
 aggtgccaa aacatacttt ggggaaaaga taactctctt aataaatggt gctggaggaa 2940
 ctggatatcc atatgcaaaa taacaatact agaactctgt ctctcaccat atacaaaagc 3000
 aaatcaaaat ggatgaaagg cttaaatcta aaacctcaaa ctttgcaact actaaaagaa 3060
 50 aacaccggag aaactctcca ggacattgga gtgggcaaag acttcttgag taattccctg 3120
 caggcacagg caaccaaagc aaaaacagac aaatgggac atatcaagtt aaaaagctt 3180
 tggccagcaa aggaacaat caacaaagag aagagacaac ccacagaatg ggagaatata 3240
 tttgcaaaat attcatctaa caaggaatta ataacagta tatataagga gctcaaaacta 3300
 ctctataaga aaaaacacta ataaagctgat tttcaaaaat aagcaaaaaga tctgggtaga 3360
 cattttctca aataagtcac acaaatggca aacaggcatc tgaatatgtg ctcaacacca 3420
 55 ctgatcatca gagaatgca aatcaaaact actatgagag atcatctcat ccagttaaa 3480
 atggtcttta ttcaaaagac aggcataaac aaatgccagt gaggatgtgg ataaaaggaa 3540
 acccttggac actgttgggt ggaatggaaa ttgctaccac tatggagaac agtttgaag 3600
 ttctcaaaa aactaaaaat aaagctacca tacagcaatc ccattgctag gtatatactc 3660
 caaaaaaggg aatcagtgta tcaacaagct atctccactc ccacatttac tgcagcactg 3720
 60 ttcatagcag ccaaggtttg gaagcaacct cagtgtccat caacagacga atggaaaaag 3780
 aaaaatgtgt gcacatacac aatggagtac tacgcagcca taaaaaagaa tgagatcctg 3840
 tcagttgcaa cagcatgggg ggcactggtc agtatgttaa gtgaaataag ccaggcacag 3900
 aaagacaaac ttttcatgtt ctcccttact tgtgggagca aaaaataaaa caattgacat 3960
 agaaaatagag gagaatgggt gttctagagg ggtgggggac aggttgacta ggtcaacaa 4020
 65 taatttattg tatgttttaa aataactaaa agagtataat tgggttgttt gtaacacaaa 4080

5 gaaaggataa atgcttgaag gtgacagata ccccatcttac cctgatgtga ttattacaca 4140
 ttgtatgcct gtatcaaaat atctcatgta tgctatagat ataaacccta ctatattaaa 4200
 aattaaaaat ttaattggcca ggcacggtgg ctcatgtccg taatcccagc actttgggag 4260
 gccgaggcgg gtggatcacc tgagggtcagg agtttgaac cagtctggcc accatgatga 4320
 aacctgtct ctactaaaaga tacaaaaat agccaggcgt ggtggcacat acctgtagtc 4380
 ccaactactc aggaggctga gacaggagaa ttgcttgaac ctgggaggcg gaggttgagc 4440
 tgagccgaga tcatgccact gcactgcagc ctgggtgaca gagcaagact ccatctcaaa 4500
 acaaaaacaa aaaaaagaag attaaaaatg taatttttat gtaccgtata aatataact 4560
 ctactatatt agaagttaaa aattaaaaaca attataaaag gtaattaac acttaacta 4620
 10 aaataagaac aatgtatgtg ggggttctag cttctgaaga agtaaaagt atggccacga 4680
 tggcagaaat gtgaggaggg aacagtggaa gttactgttg tttagcgtc atactctctg 4740
 taagtgcatt aattttaacc aaagacagggc tgggagaagt taaaggagca ttctataagc 4800
 cctaaaacaa ctgctaataa tgggtgaaagg taatctctat taattaccaa taattacaga 4860
 tatctctaaa atcgagctgc agaattggca cgtctgatca caccgtcctc tcattcacgg 4920
 15 tgcttttttt cttgtgtgct tggagatttt cgattgtgtg ttcgtgtttg gttaaactta 4980
 atctgtatga atcctgaaac gaaaaatggt ggtgatttcc tccagaagaa tttaggtacc 5040
 tggcaggaag cagggtggctc tgtggacctg agccacttca atcttcaagg gtctctggcc 5100
 aagaccagg tgcaaggcag aggcctgatg acccgaggac aggaaagctc ggtatgggaa 5160
 gggcgatgag aagcctgcct cggtggtgag cagcgcatga agtgccctta ttacgcttt 5220
 20 gcaagattg ctctggatag catctggaaa agcgggccag cgggaatgca aggagtcaga 5280
 agcctcctgc tcaaacccag gccagcagct atggcgccca cccggcgctg tgccagaggg 5340
 agaggagcca aggcacctcg aagtatggct taaatctttt ttccacctga agcagtgacc 5400
 aaggtgtatt ctgagggaag cttgagttag gtgcctctct taaaacagaa agtcacgga 5460
 gcacctctct caaggaaaa ccagacgccc gctctgcggt catttacctc ttctctctct 5520
 25 cctctctctg cctcgcggtg ttctgatcgg gacagagtga ccccgctgga gcttctccga 5580
 gcccggtgctg aggacctctc tgcaaaaggc tccacagacc cccgccttg agagaggagt 5640
 ctgagcctgg cttaataaca aactgggatg tggctggggg cggacagcga cggcgggatt 5700
 caaagactta attccatgag taaattcaac cttccacat ccgaatggat ttggatttta 5760
 30 tcttaatat ttcttaaat tcatcaata acattcagga ctgcagaaat ccaaggcgt 5820
 aaaacaggaa ctgagctatg ttgccaagg tccaaggact taataacat gttcagaggg 5880
 atttttcgcc ctaagtactt ttatttggtt ttcataagg gtcttagggg gcaaggga 5940
 gtacacgagg agaggcctg gcggcaggc tatgagcac gcagggccac cggggagaga 6000
 gtccccggcc tgggaggctg acagcaggac cactgacctg cctccctggg agctgccaca 6060
 35 ttgggcaacg cgaaggcggc cacgctcggt gtgactcagg accccatacc ggcttctctg 6120
 gccccccac actaaccag gaagtacagg agctctgaac cgtggaaac gaacatgacc 6180
 cttgccctgc tgcctccctg ggtgggtcaa gggtaatgaa gtggtgtgca ggaacggcc 6240
 atgtaaatca cagactctg ctgatgggga cgttctctc catcattatt catctcacc 6300
 cccaaggact gaattattcc agcaactctc tgggtgtgca caagccatga caaaactcag 6360
 40 tacaacacc actcttttac tagccacca gagcacgsc cacaccctg atatatag 6420
 agtccaggag agataggct gctttcagcc accaggctgg ggtgacaaca gcggctgaac 6480
 agtctgttcc tctagactag tagacctgg caggcactcc cccagattct agggcctgg 6540
 tgctgttccc cgaggcgccc atctgccctg gagactcagc ctgggtgccc acactgaggc 6600
 cagccctgtc tccacacct cgccctcag gcctcagctt ctccagcagc ttcttaaac 6660
 45 ctgggtgggc cgtgttccag cgctactgtc tcacctgtcc cactgtgtct tgtctcagc 6720
 acgtagctcg caggttctct cctcacatgg ggtgtctgtc tccttcccca acactcacat 6780
 gcgttgaggg gaggagattc tgcgcctccc agactggctc ctctgagcct gaaacctggc 6840
 cgtggcccc gatgcagggt cctggcgctc ggctgcagc tgacctccat ttccaggcgc 6900
 tccccgtctc ctgtcatctg ccggggcctg ccggtgtgtt cttctgtttc tgtgtctct 6960
 50 tccacgtcca gctgcgtgtg tctctgcccg ctagggtctc ggggttttta taggcatagg 7020
 acggggcggt ggtgggccag ggctctcttg ggaatgcaa catttggtg tgaaagttag 7080
 agtgctgtc ctccactagg tccacgggca caggcctggg gatggagccc ccgccaggga 7140
 cccgcccttc tctgccagc actttcctgc cccctccct ctggaacaca gagggtcagt 7200
 ttccacaagc actaagcctc ctcttcccaa aagaccagc attggcacc ctggacattt 7260
 55 gccccacagc cctgggaatt cacgtgacta cgcacatcat gtacacactc ccgtccacga 7320
 ccgaccccc ctgttttatt ttaatagcta caaagcaggg aatccctgc taaaatgtcc 7380
 tttaaacaa tggttaaaca aacgggtcca tccgcacggg ggacagtcc tcacagtga 7440
 gaggaacatg ccgtttataa agcctgcagg catctcaagg gaattacgct gagtcaaac 7500
 tgccacctcc atgggatacg tacgcaacat gctcaaaaag aaagaatttc accccatggc 7560
 60 aggggagtggt ttagggggt taaggacggt gggggcgga gctgggggt actgcacgca 7620
 ccttttacta aagccagttt cctggttctg atggtattgg ctcagtattg ggagactaac 7680
 cataggggag tggggatggg ggaaccgga ggctgtgcca tcttgccat gccgagtg 7740
 cctgggcagg ataattgctc agagatgccc acgtcctgat tccccaaac ctgtggacag 7800
 aaccccccg gccccagggc ctttgaggt gtgatctccg tgaggacct gaggtctggg 7860
 65 atccttcggg actactcgca ggcggaaaa ggttctggga agaggcgggc 7920
 aggagggtca gaggggggca gcctcaggac gatggaggca gtcagctga ggtgaaaag 7980

5 ggagggaggg cctcgagccc aggcctgcaa gcgcctccag aagctggaaa aagcggggaa 8040
 gggaccctcc acggagcctg cagcaggaag gcacggctgg cccttagccc accagggccc 8100
 atcgtggacc tccggcctcc gtgccatagg agggcactcg cgtgcccctt ctgcatgaa 8160
 gtgtgtgggg atttgagaa gcaacaggaa acccatgcac tgtgaatcta ggattatttc 8220
 5 aaaacaaagg ttacagaaa catccaagga cagggctgaa gtgcctccgg gcaagggcag 8280
 ggcaggcacg agtgatttta ttagctatt ttattttatt tacttacttt ctgagacaga 8340
 gttatgctct tgttgcccag gctggagtgc agcggcatga tcttggctca ctgcaacctc 8400
 cgtctcctgg gttcaagcaa ttctcgtgcc tcagcctccc aagtagctgg gatttcaggc 8460
 gtgcaccacc acaccggctt aattttgtat ttttagtaga gatgggcttt caccatgttg 8520
 10 gtcaagctga tctcaaatc ctgacctcag tggatccgcc caccctcagcc tcccaagtg 8580
 ctgggattac aggcattgag cactgcacct ggctattta accattttaa aacttccctg 8640
 ggctcaagtc acaccactg gtaaggagt catggagttc aatttcccc ttactcagga 8700
 gttaccctcc tttgatatt tctgtaattc ttcgtagact ggggatacac cgtctcttga 8760
 cataattaca gtttctgtga ccacctgtta tccccggga cccactgcag gggcagctgg 8820
 15 gaggctgcag gcttcaggtc ccagtggggt tggcatctgc cagtagaaac ctgatgtaga 8880
 atcagggcgc aagtggtgac actgtcctga atctcaatgt ctcagtgtgt gctgaaacat 8940
 ctgagaaatta aagtcctacc ctccactct actgggattg agccccctcc ctatcccccc 9000
 ccaggggag agggagttcct ctactcctg tggagggaagg aatgatactt tgttattttt 9060
 cactgctggt actgaatcca ctgtttcatt tgttggtttg tttgttttgt tttgagaggg 9120
 20 ggtttcactc ttgttgctca ggctggaggg agtgcaatgg cgcgactctg gcttacttga 9180
 gcctctgctc cccaggttca agtgattctc ctgcttccgc ctccccattg cgtgggatta 9240
 caggcacccc ccaccatgcc cagctaattt tttgtatttt tagtagagac ggggggtggg 9300
 ggggttcacc atgttgccca ggctggtctc gaacttctga cctcagatga tccacctgcc 9360
 tctgctcctc aaagtgtgg gattacaggt gtgagccacc atgcccagct cagaatttac 9420
 25 tctgtttaga aacatctggg tctgagtag gaagctcacc ccactcaagt gttgtggtgt 9480
 ttttaagcaa tgatagaatt tttttattgt tgttagaaca ctcttgatgt tttacactgt 9540
 gatgactaag acatcatcag cttttcaaa acacactaac tgcacccata atactgggg 9600
 gtcttctggg tatcagcaat cttcattgaa tgcggggagg cgtttctctg ccactgacat 9660
 gggtttaatt actccagcat aatcttctgc tttccattct tctcttccct cttttaaaat 9720
 30 tgtgttttct atgttggctt ctctgcagag aaccagtgtg agctacaact taacttttgt 9780
 tggaaacaaat tttccaaacc gcccctttgc cctagtggca gagacaattc acaaacacag 9840
 ccccttaaaa aggccttaggg atcactaagg ggatttctag aagagcgacc tgtaatccta 9900
 agtatttaca agacgaggtt aacctccagc gagcgtgaca gcccaggagg ggtgcgaggc 9960
 35 ctgttcaaat gctagctcca taaataaagg aatttccctc ggaggtttct gaaagtagga 10020
 aagggttaca ttaagggttc gtttgttagc atttcagtgt ttgccgacct cagctacagc 10080
 atccctgcaa ggcctcgga gacccagaag tttctcgtcc ccttagatcc aaacttgagc 10140
 aaccggaggt ctggattcct ggggaagtcct cagctgtcct gcggttctgc cggggcccca 10200
 ggtctggagg ggaccagtg ccgtgtggct tctactgctg ggctggaagt cgggctcctc 10260
 40 agctctgcag tccgaggtt ggagccaggt gcctggaccc cgaggctgcc ctccacctg 10320
 tgcgggctgg atgtgaccag atgttggcct catctgccag acagagtgc gggggccagg 10380
 gtcaaggccg ttgtggctgg tgtgagcgc ccggtgcgc gccagcagga gcgcctggct 10440
 ccatttccca ccttttctc acgggaccgc cccggtgggt gattaacaga tttgggggtg 10500
 tttgtcatg gtggggaccc ctgcgcctc gagaacctgc aaagagaaat gacgggctg 10560
 tgtcaaggag cccaagtgc ggggaagtgt tgcaggagg cactccggga ggtcccgct 10620
 45 gcccgtccag ggagcaatgc gtccctgggt tgcgtcccag ccgcgtctac gcgcctccgt 10680
 cctccccctc acgtccggca ttcgtgtggt ccggagcccc acgccccgcg tccggacctg 10740
 gaggcagccc tgggtctccg gatcaggcca gcggccaaag ggtcgccgca cgcacctgt 10800
 cccagggcct ccacatcatg gcccctccct cgggttaccc cacagcctag gccgattcga 10860
 cctctctcgg ctggggccct cgttggcgtc cctgcacctt gggagcgcca gcggcgcgcg 10920
 50 ggcggggaag cgcggccca acccccggtt ccgcccggag cagctgctgt gtcggggcca 10980
 ggccgggctc ccagtggatt cgcgggcaca gacgcccagg accgcgtcc ccactggcg 11040
 gagggactgg ggaccgggc acccgtcctg ccccttcacc ttccagctcc gcctcctccg 11100
 gcgggacccc gcccgtccc gacccctccc ggggtccccg cccagcccc tccgggcccc 11160
 cccagcccc ccccttctt tccggggccc cgcctctcc tcggggcgcg agtttcaggc 11220
 55 agcgtgctg cctgctgcgc acgtgggaag ccctggcccc ggccaccccc gcgatg 11276

<210> 4

<211> 104

<212> DNA

60 <213> Homo sapiens

<400> 4

65 gtgggctccc ccggggtcgg cgtccggctg ggggtgaggg cggccggggg gaaccagcga 60
 catgaggaga gcagcgaggg cgactcaggg cgcttcccc gcag 104

10

15

20

25

30

35

40

43

5

5

6

6

ctctagtaat ctagtaaatc tttttttaa ttgctcttag tactgccaca ctgggcttct 3600
 tttgattagt attttctctgc tgtgtctggt ttctgccttt aatttatata tatatatata 3660
 tttttttttt ttttgagaca gagtcttggc ctgtcgtcca ggggtgagtc agtgggtgta 3720
 5 tcacaggcca gtgtaacttt taccttctgg cctgagccgt cctctcacct cagcctcctg 3780
 agtagctgga actgcagaca cgcaccgcta cacttggtta atttttaaat tttttctgga 3840
 gacagggtct tgcgtgtgtg cccaggctgg tctcaaactc ttggactcaa gggatccatc 3900
 tacctcggct tcccaaagtg ctgaattaca ggcagtagcc accatgtctg gcctaatttt 3960
 caacactttt atattcttat agtgtgggta tgcctgttta acagcatgta ggtgaatttc 4020
 10 caatccagtc tgacagtcgt tgtttaactg gataacctga tttattttca tttttttgtc 4080
 actagagacc cgcctgggtg actctgattc tccacttgcc tgttgcatgt cctcgttccc 4140
 ttgtttctca ccacctcttg ggttgccatg tgcgtttcct gccgagtggt tgttgatcct 4200
 ctctgttgcct cctgggtcact gggcatttgc ttttatttct ctttgcttag tgttaccctc 4260
 tgatcttttt attgtcgttg tttgcttttg tttattgaga cagtctcact ctgtcaccga 4320
 ggcctggagt taatggacaca atctcggtc actgcaacct ctgcctcctc ggttcaagca 4380
 15 gtctctcatt ctcaacctca tgagtagctg ggattacagg cggccaccac cagcctggc 4440
 taatttttgt atttttagta gagataggct ttaccatgt tggcagggt ggtctcaaac 4500
 tctgacctc aagtgtctg cccgcttgg cctccacag tgcctgggt acaggtgcaa 4560
 gccacctg cggcatacc ttgatctttt aaaatgaagt ctgaaacatt gctacccttg 4620
 20 tccctgagcaa taagacctt agtgtatttt agctctggcc acccccagc ctgtgtgtgt 4680
 ttttccctgc tgacttagtt ctatctcagg catcttgaca ccccacaag ctaagcatta 4740
 ttaatatgtt tttcgtgtt gagtgttct gtacttttg ccccgccctc ttttccctcc 4800
 tttgttcccc gtctgtcttc tgtctcaggc ccgctgtctg ggggtccctt ccttgtctct 4860
 tgcgtgtgtc ttctgtcttg ttattgtctg taaacccag ctttacctgt gctggcctcc 4920
 atggcatcta ggcagctccg gggacctctg cttatgatgc acagatgaag atgtggagac 4980
 25 tcacgaggag ggcgtgtcat ttggccctg agtgtctgga gcaccacgtg gccagcgttc 5040
 cttagccagt gagtgtacag aacgtccgct cgcctcgggt tcagcctgga aaacccagg 5100
 catgtcgggg tctgtgtggt ccgctgtgtc gagtgtgaaa tcggtgcaa cctggtgtgt 5160
 ggcctcagctc tgacgtgtgt cctgtgctgg gtagtgtctg cttcctccct tctgtctggg 5220
 aaccaggaca aaggatgagg ctccgagccg ttgtcgccca acaggagcat gacgtgagcc 5280
 30 atgtggataa ttttaaaatt tctaggctgg ggcgtgtggc tcacgcctgt aatcccagca 5340
 ctttgggagg ccaaggcggg tggatcacga ggtcaggagg tcgagaccat cctggccaac 5400
 atgatgaac cccatctgta ctaaaaacac aaaaattagc tgggctgtgt ggcgggtgcc 5460
 tgtaatccca gctactcggg aggtctgagg aggagaattg cttgaacctg ggagtgtgaa 5520
 gttgcagtga gccgacattg caccactgca ctccagcctg gcaacacagc gagactctgt 5580
 35 ctcaaaaaaa aaaaaaaa aaaaaaaa aattctagta gccacattaa aaaagttaaa 5640
 aagaaaaggt gaaattaatg taataataga ttttactgaa gccagcatg tccacacctc 5700
 atcatttttag ggtgttatgt gtgggagcat cactcacagg acatttgaca ttttttgagc 5760
 tttgtctgcg ggtacccgtg tgtaggtccc gtgcgtggcc atctcggcct ggaacctgct 5820
 40 ggcctcccat ggccatggct gttgtaccag atggtgcagg tccgggatga ggtcgccagg 5880
 ccctcagtga gctggatgtg cagtgtccgg atggtgcacg tctgggatga ggtcgccagg 5940
 ccctgctgtg agctggatgt gtggtgtctg gatggtgcag gtcagggtgt aggtctccag 6000
 gccctcgggt agctggaggt atggagtcag gatgatgcag gtcgggggt aggtcgccag 6060
 gccctgctgt gagctggatg tgtggtgtct ggatgggtga ggtcagggtg gaggtctcca 6120
 45 ggcctcgggt aagctggagg tatggagtcc ggatgatgca ggtccgggtg gaggtcgcca 6180
 ggcctctgtg tgagctggat gtgtggtgtc tggatggtgc aggtctgggt tgaggtcacc 6240
 aggcctctgc gtgagctggg tgtgcgtgtg ctggatggtg cagggtctga gtaggtctgc 6300
 cagacggtgc cagacctgc ggtgagctgg atatgcgggt tccggatggt gcaggtctgg 6360
 ggtgaggttg ccaggccctg ctgtgagtgt gatgtgggtg gtcgggatgc tgcaggctcc 6420
 50 gtgtgaggtc accaggccct gctgtgagct ggatgtgtgg tgtctggatg gtgcaggctc 6480
 ggggtgaagg tcgccaggcc cctgcttctg agctggatgt gttgtgtctg gatgtgtgg 6540
 gtctggagtg aggtcgccag gccctcgggt agctggatgt gcagtgtcca gatgtgtcag 6600
 gtccgggtgt aggtcgccag accctgcggg gagctggatg tgcgtgtctt ggtgtgtgca 6660
 ggtctggagt gaggtcgcca ggcctcgggt gagctggatg tatggagtcc ggtgtgtgcc 6720
 55 ggtccgggtg gaggtcgcca gacctgtctg tgagctggat gtgcgtgtgc tggatggtac 6780
 aggtctggag tgaggtcgcc agacctgtct gtgagctgga tatgcgggtg ccggatggtg 6840
 caggctcagg gtgaggtctc caggccctgc gtgagctgga ggtatggagt ccggatgatg 6900
 caggctcagg gtgaggtctc caggccctgc tgtgaactgg atgtgcggcg tctggatggt 6960
 gcaggtctgg ggtgtgtgtc ccaggccctc ggtgagctgg aggtatggag tccggatgat 7020
 60 gcaggtctgg ggtgaggtcg ccaggccctc ctgtgagctg gatgtgcggc gctctggatg 7080
 tgacagctct ggtgtgtgtc ccaggccctc cgtgtgagct gaggtatgga gtcgggatga 7140
 tgacagctcc ggtgtgaggt gccaggccct gctgtgagct ggtgtgtctg tatccggatg 7200
 gtgcagctcc ggtgtgaggt gccaggccct gctgtgagct ggtgtgtctg tatccggatg 7260
 gtgcaggtct ggggtgaggt caccaggccc tgcgggtgagc tgggtgtgtc ggtgtccggt 7320
 65 gctgcaggtc cgggtgaggt tgcggaggtc ctgcgtgagc tggatgtgtc ggtgtccggt 7380
 gtcggatggt tgcaggtcca ggtgtgaggt gctaggccct tgggtgggtc gatgtgtcgt 7440

gtccggatgg tgcaggctctg ggggtgaggtc gccaggccctt tggtagagctg gatgtgagg 7500
 gtctgcatgg tgcaggctctg ggggtgaggtc gccaggccctt tggtagagctg gatgtgagg 7560
 gtccggatgg tgcaggctccg gcgtgaggtc gccaggccctt gctgtgagct ggatgtgagg 7620
 tgcctggatg gtgcagggtcc ggggtgaggt agccaaggcc ttcggtagagc tggatgtggg 7680
 5 gtgtccggat ggtgcaggtc cggggtgagg tgcaggccctt ctgaggtag ctgatatgc 7740
 ggtgtccgga tgggtgcaggc ccggggtgag gtccaggcc cctgaggtag gctggatgtg 7800
 cgggtgtctgg atggtgcagg tccggggtga ggtcgccagg cctgctgtg agctggatgt 7860
 gctgtatccg gatggtgcag gtccggggtg aggtcgccag gccctgcagt gagctggatg 7920
 10 tgccttatcc ggatggtgca ggtctggcgt gaggtcgcca ggccctgcgg ttagctggat 7980
 atgagggtgtc ggatggtgca ggtccggggt gaggtcacca ggccctgcgg ttagctggat 8040
 gtgcgggtgtc cggatggtgc aggtctgggg tgagggtcgc aggccctgct gtgagctgga 8100
 tgtgctgtat ccggatggtg cagggtccggg gtgagggtcgc caggccctgc ggtgagctgg 8160
 atgtgctgta tccggatggt gcagggtctg cgtgagggtc caggccctgc cgtgagctg 8220
 gatgtgcagt gtacggatgg tgcaggctccg ggggtgaggtc gccaggccctt gcgggtggg 8280
 15 gtatgtgtgt tgcctggatg gtgcagggtcc ggggtgaggt cgccaggccc tgcggtagagc 8340
 tggatgtgtg gtgtctggat gctgcaggtc cggggtgagt tgcaggccctt ctggtgagc 8400
 tggatatgcg gtgtcccggt gtccgaatgg tgcagggtcca gggtagaggt gccaggccctt 8460
 tgggtgggctg gatgtgagggt gtccggatgg tgcagggtcgc gggtagaggt gccaggccctt 8520
 20 tggtagagctg gatgtgagggt gtccggatgg tgcagggtcgc gggtagaggt gccaggccctt 8580
 cggtagatctg gatgtggcat gtccctctcg ttttaag 8616

<210> 6
 <211> 2089
 <212> DNA
 25 <213> Homo sapiens

<400> 6
 gtactgtatc cccacgccag gcctctgctt ctccaagtcc tggaaaccca gcccgccctc 60
 agcatgcgcc tgtctccact tgcctgtgct tccctggctg tgcagctctg ggcctgggagc 120
 30 caggggcccc gtccacaggcc tgggtccaagt ggattctctg caaggctctg actgcctgga 180
 gctcacgttc tcttacttct aaaatcagga gtttgtgcca agtggctctt aggggtttgta 240
 aagcagaagg gatttaaat agatggaaac actaccacta gcctccttgc ctctccctgg 300
 gatgtgggtc tgattctctc tctctttttt tttctttttt tgagatggag tctcactctg 360
 ttgcccaggc tggagtgcag tggcataatc ttggctcact gcaacctcca cctcctgggt 420
 35 ttaagcgatt caccagcctc agcctcctaa gtactggga ttacaggcac ctgccaccac 480
 gcctggctaa tttttgtact tttaggagag acgggggttc accatgttgg ccaggctgggt 540
 ctccgaactca tgacctcagg tgatccacc accttgccct cccaaagtgc tgggtttaca 600
 ggctaagcca ccgtgccag ccccgatcc tcttttaatt catgtgttcc tgtatgaatc 660
 40 tccaactat tgatttagg tcatgagagg ataaaaatccc acccacttgg cgactcactc 720
 cagggagcac ctgtgcaggg agcacctggg gataggagag tccaccatg agctaacttc 780
 taggtggctg catctgaatg gctgtgagat tttgtctgca atgttcgggt gatgagagt 840
 tgagattgtg acagattcaa gctggatttg catcagttag ggacgggagc gctggctctg 900
 gagatgccag cctggcttag cccaggccat ggtattagct tctccgtgct ccccccaggc 960
 tgactgtgga gggctttagt cagaagatca gggcttcccc agctccctctg cacactcgag 1020
 45 tccctggggg gccttgtgac acccatgccc ccaaatcagg atgtctgcag agggagctgg 1080
 cagcagacct cgtcagagg aacacagcct ctgggctggg gaccccgagc tgggtctggg 1140
 gccatttctc tgcattctgg ggagggtcag ggccttccct gtgggaacaa gttataacac 1200
 aatgcacctt acttagactt tacacgtatt taatgggtgt cgacccaaca tggtcatttg 1260
 accagtattt tggaaagaat ttaattgggg tgaccggaag gagcagacag acgtgggtgg 1320
 50 ccccaagatg ctctctgtca ctactgggac tgtgttctg cctggggggg cctggaggcc 1380
 cctcctcccc ggacagggtc ccgtgccttt tctactctgc tgggctctgc gctgctggtc 1440
 agggcaccag ctccggagca cccgcggccc cagtgtccac ggagtgcag gctgtcagcc 1500
 acagatgccc aggtccaggt gtggccgctc cagccccctg gcccccattg gtgggttttg 1560
 gggaaaaggc caagggcaga ggtgtcagga gactgggtgg ctcattgagag ctgattctgc 1620
 55 tcttggctg agctgcctg agcagcctc cccgccccct ccatctgaag ggatgtggct 1680
 ctttctacct ggggggtcct cctggggcca gccttgggt accccagtgg ctgtaccaga 1740
 gggacaggca tctgtgtggt aggggcatgg gttcacgtgg ccccatagtc agcctgggac 1800
 caggctcccc ggtgctgatg gtgggacagt caccctgggg gttgaccgcc ggactggggc 1860
 tccccagggt tgactatagg accaggtgtc cagggtgcct gcaagtagag gggctctcag 1920
 60 aggcgtctg ctggcatggg tggacgtggc cccgggcatg gccttcagcg tgtgctgccg 1980
 tgggtgcccc gagccctcac tgaagctggg ggggcttctg gcttccccctg agcttcccc 2040
 tagtctgttg tctggcttag caagcctcct gaggggctct ctattgcag 2089

<210> 7
<211> 687
<212> DNA
<213> Homo sapiens

5
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079
1080
1081
1082
1083
1084
1085
1086
1087
1088
1089
1090
1091
1092
1093
1094
1095
1096
1097
1098
1099
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1180
1181
1182
1183
1184
1185
1186
1187
1188
1189
1190
1191
1192
1193
1194
1195
1196
1197
1198
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1240
1241
1242
1243
1244
1245
1246
1247
1248
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1270
1271
1272
1273
1274
1275
1276
1277
1278
1279
1280
1281
1282
1283
1284
1285
1286
1287
1288
1289
1290
1291
1292
1293
1294
1295
1296
1297
1298
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1310
1311
1312
1313
1314
1315
1316
1317
1318
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368
1369
1370
1371
1372
1373
1374
1375
1376
1377
1378
1379
1380
1381
1382
1383
1384
1385
1386
1387
1388
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1489
1490
1491
1492
1493
1494
1495
1496
1497
1498
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1510
1511
1512
1513
1514
1515
1516
1517
1518
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564
1565
1566
1567
1568
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1589
1590
1591
1592
1593
1594
1595
1596
1597
1598
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1670
1671
1672
1673
1674
1675
1676
1677
1678
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1689
1690
1691
1692
1693
1694
1695
1696
1697
1698
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889
1890
1891
1892
1893
1894
1895
1896
1897
1898
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1910
1911
1912
1913
1914
1915
1916
1917
1918
1919
1920
1921
1922
1923
1924
1925
1926
1927
1928
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
2032
2033
2034
2035
2036
2037
2038
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2050
2051
2052
2053
2054
2055
2056
2057
2058
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2070
2071
2072
2073
2074
2075
2076
2077
2078
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2099
2100
2101
2102
2103
2104
2105
2106
2107
2108
2109
2110
2111
2112
2113
2114
2115
2116
2117
2118
2119
2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159
2160
2161
2162
2163
2164
2165
2166
2167
2168
2169
2170
2171
2172
2173
2174
2175
2176
2177
2178
2179
2180
2181
2182
2183
2184
2185
2186
2187
2188
2189
2190
2191
2192
2193
2194
2195
2196
2197
2198
2199
2200
2201
2202
2203
2204
2205
2206
2207
2208
2209
2210
2211
2212
2213
2214
2215
2216
2217
221

agctttattg aggagacat atcttccctt gaactatggt cgggtttata gtaagtcagg 180
 ggtgtggagg cctcccctgg gctcccctgt ctgtttcttc cactctgggg tegtgtgggt 240
 cctgctgtgg tgtgtggcgg gtgggcaggg ctccaggcc tcttctgtgt cattggcctg 300
 gatgtggccc tggctacgct ccgtccctgg aattcccctg cgagttggag gctttctttc 360
 5 tttctttttt tctttctttt tttttttttt tgataacaga gtctcgctct tttttgcca 420
 ggtctggagt gtttggcgtg atcttggctc actgcaacct gtgtcttctg agttcaagca 480
 attctcttgc ctacgctcc caagtagctg gaattatagg cccccaccac catgctgact 540
 aatttttgta atttttagtag agacgagggt tctccatgtt gggcaggctg gtctcgaact 600
 cctgacctca ggtgatcctc ccacctcggc ctcccaagt gctgggatga cagggtgtaa 660
 10 ccgcccgcgc cggccgagac tcgcttccct cagcttccgt gagatctgca gcgatagctg 720
 cctgcagcct tgggtgctgac aacctccgtt tcccttctcc aggtctcgct aggggtcttt 780
 ccatttcatg actctcttca cagaagagtt tcacgtgtgc tgatttccc gctgttctct 840
 gcgtaatggt tgtctgctgt ttatcgatgg cctccttcca tttcttttag gctttgttta 900
 ttgttgtttt tccggctcct tgaaggaaaa gtttcgatta tggatgtttg aacttctctt 960
 15 tctaaacaag catctgaagt tgcggttttc cctctaaagc agggatcccg agggccctgg 1020
 ctgtggagtg gcaccggtct ggggcctgtt aggaaccccg cgcacagcgg gaggctaggt 1080
 ggggtgtggg gagccagcgt tccgcctga gcccgcctc tctcagatca gcagtggcat 1140
 gcggtgctca gaggcgcaca caccctactg agaactgtgc gtgagagggg tctagatctc 1200
 gtgctcctta tgggaatcta atgcctgatg atctgaggtg gaaccgtttg ctcccaaac 1260
 20 catcccttcc cccactgctg tctgtggaa aaatcgctct ccacgaacc agtccctggt 1320
 accacaatgg ttggggacc tgtgctaaa acctgcttca gcagcctctc gtcagtgttg 1380
 atattattgg tttctgtgtg tgagtccaga ataattacgg atttctgtga tgccttccgc 1440
 cgacctcaga cccatgggct atttgtgggc gtgttgctct ctcctgggtt gggaaagggt 1500
 caggcccat gtaccttctt gttactgcct tccagggttg. tcttcagggt tgaatcgta 1560
 25 tccgtgtggg tttagccac ggccttgcgc ccagctcctg ggggctgggg aacatgctga 1620
 agcacagagt caccgtgcgc gtcttttgat gcttcacaag ctccaggcct cctgtgtccg 1680
 tgttagtggt tgtcacgtgc ctgctccat cctgtcttgg ggacgcaggg gcttagcagg 1740
 tcccgtagta aatgacaagc gtccgtgggg agtctgcaga ataggaggtg ggggtgccc 1800
 tctctctccc gctcttccag actcttctcc tgcctgtgct gtggctgcac ctgcatccct 1860
 30 gcaatccctc cagcactggg ctggagaggc ccgggagctc gactgtccact tgtgccactg 1920
 gactgtggat ggcagtcggt caccgggggtc tgatgtgtgg tgactgtgga tggcggttgg 1980
 tccaggggt ctgatgtgtg gtgactgtgg atggcggtcg tggggtctga tgtgtgact 2040
 gtggatggcg gtctgtgggt ctgatgtgtg gtgactgtgg atggcggtcg tggggtctga 2100
 tgtgtgact gtggatggcg gtctgtgggt ctgatgtgtg gactgtggat ggcgtctgtg 2160
 35 ggggtctgat tgggtactgt ggtggcagt cgtggggtct gatgtgtgtg gactgtgga 2220
 ggcgtctgtg ggggtctgat tgggtactgt ggtggcagt cgtggggtct gatgtgtgtg gactgtgga 2280
 gactgtggat ggcgtctgtg ggggtctgat tgtgtgact gtggatggcg gtctgtgggt 2340
 ctgatgtgtg gtgactgtgg atggcggtcg tggggtctga tgtgtgtgtg ctgtggatgg 2400
 40 cgggtctgtg gtctgatgtg gtgactgtgg atggcggtcg tggggtctga tgtgtgtgtg ctgtggatgg 2460
 ctgtggatgg tgatcgggtc cagggggtctg atgtgtgtg actgtggatg gcggtctgtg 2520
 ggtctgatgt gtgtgtactg tggatgtgtg tgggtcacag ggtctgtatg tgtgtgtact 2580
 gtggatggcg gtctgtgggt ctgatgtgtg gtgactgtgg atggcggtcg gtcccggggg 2640
 tctgatgtgt ggtgactgtg gatggcgatc ggtcacaggg gtctgatgtg tgggtactgt 2700
 ggatggcggt cgtggggtct gatgtgtgtg gactgtggat ggcgtctgtg ggggtctgtg 2760
 45 tgtgtgtact gtggatggcg gtctgtgggt ctgatgtgtg gactgtggat ggcgtctgtg 2820
 ggggtctgat tgggtactgt ggtggcggt cgtggggtct gatgtgtgtg gactgtggat 2880
 ggcgtgtgtg cccgggggtc tgatgtgtgg tgactgtgga tggcggtcgt ggggtctgtg 2940
 gtgtgtactg tggatggcag tctgtgggtc tgatgtgtgg tgactgtgga tggcggtcgt 3000
 50 ggggtctgat gtgtgtgtgac tgtggatggc ggtcgtgggg tctgatgtgt ggtgactgtg 3060
 gatggcggtc gtggggtctg atgtgtgtgt actgtggatg ggcgtctgtg ggtctgtgtg 3120
 ggtgactgtg gatggcggtc gtggggtctg atgtgtgtgt actgtggatg gtgatcgggtc 3180
 acagggggtct gatgtgtgtg gactgtggat ggcgtctgtg ggggtctgtg tgtgtgtact 3240
 gtggatggcg gtctgtgggt ctgatgtgtg gactgtggat ggcgtctgtg ggtctgtgtg 3300
 55 tgtgtgtact gtggatggcg gtctgtgggt ctgatgtgtg gtgactgtgt atggcagtcg 3360
 gtccacagggt tctgatgtgt ggtgactgtg gatggcggtc gtggggtctg atgtgtgtgt 3420
 actgtgtgtg ggcgtctgtg ggtctgtgtg gtgtgactg tggatggcg tctgtgggtc 3480
 tgatgtgtgt tgactgtgga tggcggtcgt ggggtctgtg gtgtgtactg tggatgggtg 3540
 tccgtcacag ggtctgtgtg tgtgtgtgtg gcagggtggg tccaggtgtg gtctgtgtgt 3600
 60 actttgcgtc ctccggcccc cggccccctg tcccaaacaa gaagcttccc aggcgtctct 3660
 tgggcttcat cccgccatcg ggttggccg caggtccaca cgtcctgtatc ggaagaaaca 3720
 agtgcacagc tctggccggg gcaggccaca tttgtggctc atgcctctct cctgtccggc 3780
 ag 3782

12 / 18

<210> 11
<211> 980
<212> DNA
<213> Homo sapiens

5

<400> 11
gtctggggcac tgccttgacag gggtggggcac ggacteccag cagtgggtcc tccctggggc 60
aatcactggg ctcatgaccg gacagactgt tggccctggg gggcagtggg gggaaatgagc 120
tgtgatgggg gcatgatgag ctgtgtgcct tggcgaaatc tgagctgggc catgccaggc 180
10 tgcgacagct gctgcattca ggcacctgct cactgttgac tgcgcgccct ctctccagtt 240
ccgcagtgcc ttgtttcatg atttgtctaa tgtcttctct gccagttttg atcttgaggc 300
caaaggaaag gtgtccccct cctttaggag ggcaggccat gtctgagccg tgtcctgccc 360
agctggcccc tcatgtctgg gtctgaggcc aaaggaaacg tgtccccctt cttaggaggga 420
cgggccgtgt ttgagccacg ccccgctgag cgggcctctc agtgcctggg ctgtcccagc 480
15 ggccctgtgg ccctttgcag atgtggtctg tccacgtggc cctgtggctc tttgcagatg 540
cctgttagca ctgtctcggc tctaggggac agtgcgtgct accgcatgag gctcagagac 600
ctctggggga atttcttggt ctcccagggt ggggggtggg gtggcctggg tgcctggggc 660
ccagaccctg tgcggcgag ctgggcagca actcctggat cacatatgcc atccggggcca 720
cggtgggctg tgtgggtgtg agcccagctg gacccacagg tggcccagag gagacgttct 780
20 gtgtcacaca ctctgcctaa gcccatgtgt gtctgcagag actcggcccc gccagccccc 840
gatggccctg cattccagcc cagcccccga cttcatcaca aacactgacc ccaaaaaggga 900
cggagggtct tggccacgtg gtccctgctg tctcagcacc caccggctca ctcccatgtg 960
tctcccgctc gctttcgag 980

25

<210> 12
<211> 2485
<212> DNA
<213> Homo sapiens

30

<400> 12
gtgagtcagg tggccagggt ccattgcctt gcgggtggct gggcgggctg gcagggtctc 60
tgctcacctc tctcctgccc ctccccact gnccttctgc ccggggccac cagagtcctc 120
ttttctggcc ccgccccctt ccggctcctg ggctgcaggc tcccagggcc ccggaacatc 180
ggctcggctt ggcgcagccg gagcggagca ggtgccacac gaggcctgga aatggcaagc 240
35 ggggtgtgga gttgctcctg cgtggaggac gaggggcggg ggggtgtgtc gggtcagggt 300
tgcgcccagc gtttgagcct gcagcttgtc agctccaagt tactactgac gctggacacc 360
cggctctcac acgcttctat ctctctctcc cgatacaaaa ggattttatc cgatttctat 420
tctctgctct gtctgtgtac ccccgcgagg gcgcgggctc ttctctctgt gactagattt 480
cccatctgga aagtgcgggg ttgaccgtgt agtttgcctc tctcgggggg cctgtggttg 540
40 ccatggggca ggcggcctgg gagagctgcc gtccacacag cactgggtga gccacactca 600
cgggtgtaga gccacagtgc ctggtgccac atcacgtcct ctggatttta agtaaaacca 660
cacacctccc ggcaggcctc tgcctgcgac cctgtgtgtg cctggggaga gtggtagcac 720
ggaggaaatt cgtgcacact caaggctcct agcaaggcca tccgcagcca ggtggaacgt 780
ggaggcctct ctctgggagc gtctccagcg gataaaggac tgtgcacagc ttcgggaagct 840
45 tttattttaa aatataacta ttaattattg cattataagt aatcactaat ggtatcagca 900
attataatat ttattaaagt ataattagaa atattaaagta gtacacacgt tctggaaaaa 960
cacaaattgc acatggcagc agagtgaatt ttggccgagg gacacgtgtg cacatgtgtg 1020
taagcggccc ccaggccccc agaattcgtt gacaaagtca cctccccaga gaagccacca 1080
cgggcctcct tctgtggtct gaattttatt aagatggatc aagtcacgta ccgtccacgt 1140
50 gtggcagggc tttggggaat gtgagggtat gactgcgtcc tcatgccttg acagacagga 1200
ggtgactgtg tctgtcctgt ccttaggaca cggacaggcc cgaagctcta gtccccatcg 1260
tggtccagtt tggcctctga ataaaaacgt cttcaaaacc tgttgcccca aaaactaaga 1320
acagagagag tttcccatcc catgtgctca caggggcgta tctgcttgcg ttgactcgct 1380
gggctggccg gactcctaga gttggtgctg gtgcttctgt gcaaaaagtg cagtccctct 1440
55 gccatcact gtgatctctg caccagcaag gaaagcctct tttctttctt ctcttttttt 1500
ttttttgaga cggaacgtca ctgttgtctg cctgggcttg agtgacagtgg cgcgatctca 1560
actcactgca acctccgctt cccgggttcc agcattttct ctgctctcag ctcccagca 1620
gctgagatta caggcaccca cccctgcgc ctggctaat tttgtatttt tagtagagag 1680
gggtttttgc catgttgccc aggctggtct cgaactcctg acctcaggtg atccaccac 1740
60 ctgcgctccc caaagtgcgt ggattacagg tgtgagccat cacgccagc cggaaagcct 1800
ctttttaagg tgaccaccta tagcgtctcc gaaaaataac aggtcttgtt tttgcagtag 1860
gctgcaagcg tctcttagca acaggagtgg cgtcctgtgg gctctgggga tggctgaggg 1920
tcgctggca gccatgcctt ctgtgtgcac ctttaggttc caccgggcta tctgtctctc 1980
actgtttgtc tgaaaacgca cccttggcat cctgttttgg agagtctctg ctctcgttg 2040
65 gtcagtctga aactaggggc aaggttgtat ccgttggcgc gcagcggtca catgtagggt 2100

13 / 18

catgagtctt tcaccgtgga caaattccctt gaaaaaaaaa aaaggagtcc ggtaaagcat 2160
 tcattccggg tcaagtgtct ggttctgtga ataaactcta agattttaaga aaccttaatg 2220
 aaagaaaacc ttgatgatcc agagcaagga tgtggtcaca cctgtggctg gatctgtttc 2280
 agccgccccca gtgcatgggt agagtgggga gcagggaattg tttgttcaga ggtctcatct 2340
 5 ggtatgtttc tgaggtgttt gccggctgaa tggtagacgt gtcgtttgtg tgtatgaggt 2400
 tctgtgtctg tgtgtggctc ggtttgagtg tacgcatgtc cagcacatgc cctgcccgtc 2460
 tctcacctgt gtcttcccgc cccag 2485

<210> 13
 <211> 1984
 <212> DNA
 <213> Homo sapiens

<400> 13
 15 gtgaggcctc ctcttcccca ggggggcttg ggtgggggtt gatttgcttt tgatgcattc 60
 agtgttaata ttctctgtgc tctggagacc atgactgtct tgtcttgagg aaccagacaa 120
 ggttgtagcc ccttcttggg atgaagccgc acgggagggg ttgcacagcc tgaggactgc 180
 gggctccacg caggctctgt ccagcggcca tgtccagagg cctcagggct cagcaggcgg 240
 20 gaggggccgct gccctgcatg atgagcatgt gaattcaaca ccgaggaagc acaccagctt 300
 ctgtcacgtc acccagggtc cgttaggggt cttggggaga tggggctggt gcagcctgag 360
 gccccacatc tcccagcagg ccttcgacag gtggcctgga ctgggcgcct cttcagccca 420
 ttgcccaccc cacttgcatg ggggtctacac ccaaggacgc acacacctaa atatcgtgcc 480
 aacctaatgt ggttcaactc agctggcttt tattgacagc agttactttt ttttttttaa 540
 tacttttaagt tctaggggtac atgtgcacga cgtgcagggt agttacatat gtatacatgt 600
 25 gccatgttgg tgtgctgcac ccattaaact atcatttaca ttaggatatat ctctaatgc 660
 tatccctccc cactccccc atcccatgac agggcctggg gtgtgatgtt cccaccctg 720
 tgtccaagtg ttctcattgt tcagtcccca cctgtgagtg agaacatgtg gtgtttgggt 780
 ttctttctct gcaatagttt gctcagagtg atggtttcca gcttcgtcca tgtccctaca 840
 aaggacatga actcatcctt ttttatgact gcatagtatt ccgtgggtga tatgtgccac 900
 30 attttcttaa tccagtctat catcgatgga catttgggtt ggttgcaagt ctttgcctact 960
 gtgaatatg cgcgaataaa catacgtgtg catgtgtctt tatagcagca tgatttataa 1020
 tcctttgggt atataccag taatgggatg gctgggtcaa atggattttc tagttctaga 1080
 tccttgagga atcaccacac tgtcttccac aatgggtgaa ctagtttaca ctcccaccaa 1140
 cagtgtaaaa gtgttctggt gctggagagg atgtggacag cagttatttt tttatgaaaa 1200
 35 tagtatcact gaacaagcag acagtttagt aaggatgcgt caggaagcct gcaggccaca 1260
 cagccaattc tctcgaagac tccgggtttt tccgtgcat cttttgaaac tctagctcca 1320
 attatagcat gtacagtgga tcaaggttct tcttcattaa ggttcaagtt ctagattgaa 1380
 ataagtttat gtaacagaaa caaaaatttc ttgtacacac aacttgctct gggatttgga 1440
 ggaaagtgtc ctcgagctgg cggcacactg gtcagccctc tgggacagga tacctctggc 1500
 40 ccatggctcat ggggctgctg gcttgggctt gagggtcaca cagtgcacca tggccagctt 1560
 cctgtggata ggatctgggt ctcggatcat gctgaggacc acagctgccca tgcgtgtaaa 1620
 gggcaccacg tggctcagag ggggctgaggt tcccagcccc agctttctta cgtctctcag 1680
 ttatttttcc ctaagagtct gagaagtggg gccgcgcttg atggccttcg ttcgtcttca 1740
 gctggcacag aattgcacaa gctgatggta aacctgagt acttataatg aatgaggaat 1800
 45 tgcgttagca gtttaactgta gagagctcgt ctgttggaag gaaatttaag tttttcattt 1860
 aaccgctttg gagaatgtta ctttatttat ggctgtgtaa attgtttgac attcagctcc 1920
 tcgtagacag atactacgta aaaagtgtaa agttaacctt gctgtgtatt ttcctctatt 1980
 ttag 1984

<210> 14
 <211> 1871
 <212> DNA
 <213> Homo sapiens

<400> 14
 55 gtgaggcccc tgccgtgtgt ctgtggggac ctccacagcc tgtgggcttt gcagttgagc 60
 cccccgtgtc ctgccccctg caccgcagcg ttgtctctgc caagtcctct ctctctgccg 120
 gtgctggatc cgcaagagca gaggcgcttg gccgtgcacc caggcctggg ggccgagggg 180
 60 caccctcggg agggagtggt taccgtgcag gccctggctc tgcagagacg caccacaggt 240
 acacacgtgg tgagtgcagg cgtgacctg gctcctgctg ctctttggaa agtcaagagt 300
 gggcgctcct ggggccccag tgagaccccc aggagctgtg cacagggcct gcaggggcca 360
 ggcggcagcc tccctcccag ggtgcacctg agcctgcgga gagcaggagc tgcgtgagtg 420
 gctggccccac agcgttcgct gcggctcacgt tccctgcgtg ggttgtttgg gatcgggtgg 480
 agaatttggg tttgctgagt gctgctgtct tgaaccacgg agatgggctag gagggggttt 540
 65 cagagttgat ttttgtgaat caaactaaaa tcaggcacag gggacctggc ctacgcacag 600

gggattgtcc aatgtggtcc ccctcaaggc cgccccacag agccgggtggg cttgttttaa 660
 agtgcgattt gacgaggagc gagaacacct gaaagctgta aaggggaacc tcagaaaatg 720
 tggccgcccag ggggtggttcc aggtgctttg ctgggctgtg tttgtgaaaa cccattttga 780
 cccgcccctcc aagttccacc tccaggtcca ccctccaggc ccgcccctggg ctgggggtat 840
 5 gcttggcggt ccttgtgccc cagcccggag cacagcaggc tgtgcacatt taaatccact 900
 aagattcact cggggggagc ccaggtccca agcaactgag ggctcaggag tcctgaggct 960
 gctgagggga cagagcagac ggggaacgct gcttctgtgt ggcaagtcc tgagggtgct 1020
 ggccaggagc gtggctcaga gtgtatgtt gggtcccacc gggggcagaa ctctgtctct 1080
 gatgagtcgg cagccatgta acaggaaggc gtggccacag ggagctggga atgcaccagg 1140
 10 ggagctgcgc agctggccga ggtcccagg ccaggccaca ggaaggcag ggggacgccc 1200
 ggggcccacag cagagccgc aggaaggga ggggatgccc aggccagagc agaggctacc 1260
 gggcacaggg gggctcccgt agctgggtga gcgaggctca tgactcggcg agggaaacct 1320
 cttgacgtga agctgacgac tgggtgtgct cagctcacag cccagccagg tcccgcgctc 1380
 gagcaggaac tcagaacct cccctttgtc taaagcacag cagatgcctt caggggcatc 1440
 15 agggaaaaac agggaaagtc gttgagaaac gtcttaaaag aaggtgggat ggtggcaatt 1500
 tcttgtccag atttttagtct gcccgggacc acagatgagt ctataacggg attgtggtgt 1560
 tgccatgggg acacatgaga tggaccatca cagagccac tggggtgca cctccatct 1620
 gagtccctggc tgtcccgggt ccaggccagg tcttgcagc ctacactacc tgtcctgccc 1680
 gggagacagg gaaagcacc cgaagtctgg agcagggtcg ggtccaggct cctcagagct 1740
 20 cctgcccaggc ccagcacct gctccaaac accacttctc tgggggtttc caaagcattt 1800
 aacaagggtg tcaggttacc tctgggtga cggccccga tctgggggt gacattgccc 1860
 ctctgcctta g 1871

<210> 15
 <211> 3801
 <212> DNA
 <213> Homo sapiens

<400> 15
 30 gtgagcgac ctggccgga gtggagcctg tgcccggctg gggcaggtgc tgctgcaggc 60
 ccgttgctgc cactctgct tccgtgtggg gcaggcgact gccaatccca aagggtcaga 120
 ggccacaggc tgcccctcgt cccatctggg gctgagcaga aatgcattct tctgtgggag 180
 tgagggtgct cacaacggga gcagttttct gtgctatttt ggtaaaagga aatggtgcac 240
 35 gagacctggg tgactgagg tgtcttcaga aagcagctcg gatccgaacc caagacgccc 300
 gggccctgct gggcgtgagt ctctcaaac cgaacacagg gggcctgctg ggcagtagtc 360
 cctctgaacc cagagacctg gggccctgct gggcgtgagt ctctccgaac ccagagactt 420
 cagggccctt ttgggcgtga gtctctccgc tgtgagcccc acactccaag gctcatccac 480
 agtctacagg atgcatgag ttcattgatca cgtgtgacct atcaggggac agggccatgg 540
 40 tgtggggggg gtctctacaa aattctgggg tcttgtttcc ccagagcccc agagctcaag 600
 gccccctctc aggtctcagc acaaatgaat tgaagatgga cacagatgca gaaatctgtg 660
 ctgtttcttt tatgaataaa aagtatcaac attccaggca gggcaagggt gctcacacct 720
 ataatcccag cactttggga gggcaggtg ggtggatcac ttgaggccag gagtttgagg 780
 ccaacctaac caacatagt aaattccatt tctacttaaa aaatacaaaa attagcctgg 840
 45 cctggtggca cagcctgta gtcccgcgta tgccggaggc tgaggcagga gaattcattt 900
 aacccaggag gcagaggttg cagtgcagcc actgcactcc agcctgggca 960
 acagagttag acttcatctt aaaaaaaaa aaaaaagtat cagattccca aaaccatagt 1020
 ggacagggtt ttttttatct tgcctctcga taattttac tgggtgctgtg ctgaggccg 1080
 gaaatggggg tgccctctct tgaaggcag accttcatgg gaagagaaat aagtgggtga 1140
 50 tgggtgttaa accagaggtt taaactgggg tccgtgctgt ctgagttaac agtccagatc 1200
 tggactttgc ctctttccag aatgctccct ggggtttgct tcatggggga gcagcagggt 1260
 tggacacctc cgtgatgggg gagcagcagg tgcagacgcc ctcatgatgg gggagtgga 1320
 ggtgcagaca ccttctgcca tgggtgccag catgtccctg ttgcagctcc cccccacaa 1380
 ggaatgccgg ctctctgtgt ccccacagtc cctgcttccc tctcacagcc ttacctggct 1440
 55 ctggcctcca ctggcttgt ctgcagatt tccacatttc ctgggctccc agcacctctt 1500
 cgctctctcc aggcacctct gcagtgtctg ccataccagt cagctgtgaa ctgtccactg 1560
 cttattttgc tccccatgaa atgtattttt taggacaggc acccctggtt ccagcctctg 1620
 gcacagcatc agtgaatgtt attgaaggac aaaggacaga caaacaatc aggaaaatgg 1680
 gttctctcta aacacattgc aaagccacag aggtcagtag aggatgggtg ggcacagggt 1740
 60 catcagatgt ggggtccaat ccagaatatt ctgtgctccc aaaggccact tggtcagagt 1800
 gtgtgcttgc agaggtggct ctaaaagctc agcagtgagg gcagtggttc gccatactca 1860
 ggggtgaactc acatccctctg tgtctgaagt atacagcaga ggcttgaaag gcattctggg 1920
 gaagaaaaca ggcaaaatga ttaagaaaag tgaagaaaga aaagtggtaa gatgggaatt 1980
 tcttgtcca gatttttagtc tcccaacca cagctcagat ggtagaatgt ggtcagaact 2040
 65 gatggacaga acaatagaac aaaacggaag cctatctctc cagaacctgt tgttaattgt 2100
 gtatgtggca cagctgatgg aaaagagagt gtgtgtgtaa tttttttttc tgagaaaact 2160

		gactggaagc	aaataagttg	tgtctttaca	gcatatacca	gagcagattc	taggtagaag	2220
		aggagacaca	tgcaaacac	accagcaaca	gaaataaaac	aaaagactca	aaggggaagg	2280
		aggtgaacgt	tccctgggtt	ggtgttgggg	aaggacacac	agggagggcg	atgaaaccag	2340
		tgaggcaacg	ggcattgctt	tcactgcaga	gaaactcagc	ttgcctgagc	cacagtgaag	2400
5		atggccattc	cctggagcgt	ttgtgcacgt	gatttattta	aggcgccctg	tgaggtcctg	2460
		cacattcatc	ctctcacttt	gttctcctaa	ccacctgaga	ggtagaggag	gaaaggctcc	2520
		agggggagcag	ccgccttggg	tcacccagct	ggcaaaaggg	atgcatgatt	gcagcctggc	2580
		ctcctgctcc	ggggcccttg	ctctgcccga	ggaccccaca	caagtcagac	ccataggctc	2640
		aggggtgagcc	ggagcccag	gtcgtgttgg	ggatggctgt	gaaagaagaa	atggacgtct	2700
10		gatgcacact	tggaaggtc	ctaccagcag	cgtcaaagaa	atgcatgtga	aactgacagc	2760
		gagacccatc	cctcaaagaa	acgcacgtga	aactgatggc	gagacctgtc	cccatccctc	2820
		atgtcggctc	cttttctggg	cttgccaaga	gccagcatca	ggttgaggca	agctggaaag	2880
		acttttctgg	aaagcagctt	gtttgcattg	aagtccctac	aatgtcctgt	gtcttccctg	2940
		taattccact	tctgaagtga	ccagacatta	tcacgggtct	tatttaccat	ttccagtgtt	3000
15		ccaggcaggg	ggacttgcca	cagcaagtca	cgaacctgcc	caaatacagg	gctaaggaga	3060
		tattatgcat	cacaaaactt	gctctgccat	taaacatttt	tcaaagaatt	tttgaagaat	3120
		gttttaagtc	acaaaacgtt	tatttcaatg	tagcagtgtt	caaagctgga	tgtaaaagaa	3180
		cacaccccag	gagcctgccc	tgaatgtcat	gtgtgttcat	ctttggacat	ggacatacat	3240
		gggcagtgag	tggtggtgag	gccttgagg	acatcggtgg	gatgcctcca	tcctgcccct	3300
20		ctggagacac	catgtgtgcc	acgtgcactc	actggagccc	tgtttagctg	gtgccacctg	3360
		gctcttccat	ccctgagatt	caaacacagt	gagattcccc	acgcccact	cagtgttctc	3420
		ccacaaaaaa	cctgagtcac	acctgtgttc	actcgaggga	cgcccgagg	ccagggtccc	3480
		acagtttatt	atgtgttttt	ggctgagttc	tgtgcagatc	tcacagggc	agatgatgag	3540
		tgacaaaaa	cgcccggtgc	aggtttggat	acactcaaca	tcactagcca	ggctcctggt	3600
25		gagtttggtc	atgcagagtc	tggatggcat	gtagcatttt	gagtcattgg	agtgagcacc	3660
		cagccccctc	gggctgcagc	gcatgcccca	ggcaggacaa	ggaagcggga	ggaaggcagg	3720
		agggctcttg	gagcaagctt	tgacggaggg	ggctgggtgt	ggggcaggca	cctgtgtctg	3780
		acattccccc	ctgtgtctca	g				3801
30		<210>	16					
		<211>	880					
		<212>	DNA					
		<213>	Homo sapiens					
35		<400>	16					
		gtgagcaggc	tgatggtcag	cacagagttc	agagttcagg	aggtgtgtgc	gcaagtatgt	60
		gtgtgtgtgt	gtgcgcgcgt	gcctgcaagg	ctgatgttga	ctggctgcac	gtaagagtgc	120
		acatgtacgc	atatacacgt	gagcacatac	atgtgtgcat	gtgtgtacat	gaaggcatgg	180
		cagtggtgtg	acaggtgtgc	aagggcacaa	gtgtgtgcac	atgcgaatgc	acacctgaca	240
40		tgcatgtgtg	ttcgtgcaca	gtcgtgtggg	cattcacgtg	aggtgcatgc	gtgtgggtgt	300
		gcagtggtgag	tagcatgtgt	gcacataaca	tgtattgagg	ggtcctctgt	ttcaccccgc	360
		taggtcctca	gcaccagtgc	cactccttac	aggatgagac	gggggtcccag	gccttgggtg	420
		gctgaggctc	tgaagctgca	gccctgaggg	cattgtccca	tctgggcac	cgctgccact	480
		ccctctcctg	tgggcttctg	tgtccactcc	ccctctcctg	tgggcattta	catccactcc	540
45		actccctctc	tcctgtgggc	atcccgctcc	actccccctc	tctgtgggca	tctgcgtcca	600
		cctccccctc	ctgtgggcat	ttgcgtccac	tccctctcct	ggttccttcc	tgtcctgggc	660
		gagcctcggg	ggcaggcaga	tgacacagag	tcttgactcg	cccagggtgg	ttcgcagctg	720
		ccgggtgagg	gccaggccgg	atttactctg	gaagagggat	agtttcttgt	caaaatgttc	780
		ctctttcttg	ttccatctga	atggatgata	aagcaaaaag	taaaaactta	aaatcccaga	840
50		gagggttctc	ccgtttctca	ctcttctctg	gcgactctag			880
		<210>	17					
		<211>	3186					
		<212>	DNA					
55		<213>	Homo sapiens					
		<400>	17					
		gtgagccgcc	accaaggggt	gcaggcccag	cctccaggga	ccctccgcgc	tctgctcacc	60
		tctgaccctg	ggcttcacct	tggaaactct	gggttttagg	ggcaagggaat	gtcttacgtt	120
60		ttcagtggtg	ctgctgcctg	tgacagtttc	tgttccgcgt	gctctgtgca	aagcacctgt	180
		tctccatctc	tgggtagtgg	taggagccgg	tgtggcccca	ggtgtcccca	ctgtgcctgt	240
		gcactggccg	tgggacgtca	tggaggccat	cccaggggcag	caggggcatg	gggtaaaagag	300
		atgtttatgg	ggagctctta	cagaggaggg	tgggaagggt	tctgaacagt	agatgggaga	360
		tcagatgccc	ggaggatttg	gggtctcagc	aaagagggcc	gaggtgggtg	cagggtgagg	420
65		tcgctggccc	cacccccggg	aaggtgcagc	agagctgtgg	ctccccacac	agcccgcca	480

gcacctgtgc tctgggcatg gctgtgctcc tggaaacgttc cctgtcctgg ctggtcaggg 540
 ggtgccccctg ccaagaatcg acaactttat cacagaggga agggccaatc tgtggaggcc 600
 acaggggccag cttctgcctg gagttagggc aggtggtggc acaagcctcg gggctgtacc 660
 aaagggcagc cggggaccac agggccgggc cccacctca acaggcctcc cgagccactg 720
 5 ggagctgaat gccaggaggc cgaagccctc gcccctagag ggctgagaag gagtgtgagc 780
 atttgtgtta cccaggggcg aggtgctgcg aattaccgtg cacacttgat gtgaaatgag 840
 gtcgtcgtct atcgtggaac cccagcaagg gctcacggga gatttttcca tacaaggctc 900
 gtaccatgaa aatgggtttt aaccggagtg cttgcgcctt catgctctgg caggaggaggc 960
 10 agagccacag ctgcatgtta ccgcttttgc accagctcca gaggtctggg accaggctgt 1020
 ctcagttcca ggggtgcgtc ggctcagacc gccctcctct ctgccttctc tctctgcctc 1080
 aaatcttccc tcgtttgcat ctccttgacg cgtgcctggg ccctcgtgca agctgcttga 1140
 ctccctttccg gaaacccttg ggggtgtgct gatacaggtg ccaactgagga ctggagggtg 1200
 ctgacactgt ggttgacccc aggggtccagc tggcgtgctt ggggcctcct tggggccatga 1260
 15 tgaggtcaga ggagttttcc caggtgaaaa ctccctggga actcccaggg ccatgtgacc 1320
 tgccacctgc tctctccata ttcagctcag tcttgcctc atttcccac cagggtctct 1380
 agctccgagg agctcccgtg gagggcctgg gctcagggca gggcggtgca gtttcccac 1440
 ccatgtgggg acccttgggt agtcgcttga ttgggtagcc ctgaggaggc cgagatgcga 1500
 tggggccacgg gccgtttcca aacacagagt caggcacgtg gaaggccag gaatcccctt 1560
 20 ccctcgaggc aggagtgagg gaacggagag ctgggccccg atttcacggc agccaggctg 1620
 cagtggggca ggtgtggtg gtccacgtgg cgtcgggggc ggggtctgat tcaaatcccg 1680
 tggggctcgg ccttctcggc ccgtgctggc cgcgcctcca cacgggcttg ggggtggaccg 1740
 ccgacactct agcaggtggc tatttctccc ttggaagag agccctcac ccatgctagg 1800
 tgtttccctc ctgggtcagg agcgtggccg tgtggcaacc ccgggacctt aggccttatt 1860
 attgttttaa aaacattctg ggcctggctt ccgtgtgtgc taaatgggga aaagacatcc 1920
 25 cacctcagca gagtactga gaggctgaaa ccgggtgtgt ggcttgactg gtgtgactc 1980
 aggtcattcc agaagtggc cagggaagtca gtgagaccag gtacatgggg ggctcaggca 2040
 gtgggtgaga tgaggtacac ggggggctca ggcagtggtt gaggccagg acatgggggg 2100
 ctagggcaact ggggtgagat aggtacacgg ggggctcagg cagagggtca gaccagggtac 2160
 30 acgggggctc tgatcacacg cacatatgag cacatgtgca catgtgctgt tccatggtag 2220
 ccaggctctg gcacacctgc cccaaagtcc caggaagctg agaggccaaa gatggaggct 2280
 gacagggtcg gcgcgtgtgc tcacacctgt agtcccagca ctttgggagg ccgaggcgag 2340
 aggatccctt gagcccagga gtttaagacc agcctgagca acatagtaga accccatctc 2400
 tatgaaaaat aaaaacaaaa attagctgaa catggtgtgt tgcgcttga gttccaatac 2460
 35 ttgggaggct gaagtgggag gatcacttga gcccaggagg tggaaagctg agtgagctga 2520
 gattgcacca ctgtactgca gcctgggtga cagagtga gcccattcca acaacaacaa 2580
 agaagactga caaatgcagt ttcttggaaa gaaacattta gtaggaaactt aacctacaca 2640
 cagaagccaa gtcggtgtct cgggtgtcag gagatgagat gatgggtcct cacaccatca 2700
 cccagacccc aggggtttatg caccacaggg gcgggtggct cagaaggag gcgcaggacg 2760
 40 ttgatatacg atgacatcaa ggttgcctga cgaaggagc gattcatgat aagtaacctc 2820
 tggatacaaa ggaacaatgg ataaactgga aaccttagag gccttcccgg aacaggggct 2880
 aatcagaagc cagcatgggg ggctggcacc caggatggag ctgcttcagc ctccacatgc 2940
 gtgttcatac agatgggtga cagaaacgca gtgtacctgt gcacacacag acacgcagct 3000
 actgcacac acaagcacac acacagacat gcattgcatgc atccgtgtgt gtgcacctgt 3060
 45 gcccagagc aaacctatgc atgtgcattc atgcacgcac acaggcaccg gtggggccat 3120
 gcccacaccc acgagcaccg tctgattagg aggcctttcc tctgacgctg tccgcatcc 3180
 tctcag

<210> 18
 <211> 781
 50 <212> DNA
 <213> Homo sapiens

<400> 18
 gtatgtgcag gtgcctggcc tcagtggcag cagtgcctgc ctgctgggtg tagtgtgtca 60
 55 ggagactgag tgaatctggg ctttaggaagt tcttaccctt tttcgatca ggaagtgtt 120
 taaccacaac actgtcaggc tcgtctgccc gccctctcgt ggggtgagca gagcacctga 180
 tggaaaggac aggagctgtc tgggagctgc catccttccc accttgctct gcctggggaa 240
 gcgctggggg gcctggcttc tcctgtttgc cccatgggtg gatttggggg gcctggcctc 300
 60 tcctgtttgc cctgtggtgg gattgggctg tctcccgtcc atggcactta gggcccttgt 360
 gcaaacccag gccaaaggct taggaggagg ccaggccag gctacccac cctctcagg 420
 agcagaggcc gcgtatcac acgacagag cccgcggcgt cctctgcttc ccagtcaccg 480
 tcctctgccc ctggacactt tgtccagcat caggagggtt tctgatccgt ctgaaattca 540
 agccatgtcg aacctgcggt cctgagctta acagcttcta ctttctgttc tttctgtgtt 600

gtggaaaatt cacctggaga agccgaagaa aacattttctg tegtactcc tgcgggtgctt 660
gggtcgggac agccagagat ggagccaccc cgcagaccgt cgggtgtggg cagctttccc 720
gtgtctcctg ggaggggagc tgggtcgggc ctgtgactcc tcagcctctg ttttccccca 780
g 781

5
<210> 19
<211> 536
<212> DNA
<213> Homo sapiens

10
<400> 19
gcaagtgtgg gtggaggcca gtgcgggccc cacctgccca ggggtcatcc ttgaacgccc 60
tgtgtggggc gagcagcctc agatgtctgt gaagtgcaga cggccccggg cctgaccctg 120
ggggcctgga gccacgctcg cagccctatg tgattaaacg ctggtgtccc caggccacgg 180
15 agcctggcag ggtccccaac ttcttgaacc cctgcttccc atctcagggg cgtatggctcc 240
ccacgcttgg gagccttctg acccctgacc tgtgtcctct cacagcctct tccctggctg 300
ctgccctgag ctctctgggt cctgagcaag ttctctcccc gccccgcccgc tccagcgtca 360
ctgggctgcc tgtctgctcg ccccggtgga ggggtgtctg tcccttctact gaggttccca 420
ccagccaggg ccacaggtg caggccctgc ctgcccgcc acccacacgt cctaggaggg 480
20 ttggaggatg ccacctctgg cctcttctgg aacggagtct gatatttgcc ccgcag 536

<210> 20
<211> 3179
<212> DNA
25 <213> Homo sapiens

<400> 20
atctcatgtt tgaatcctaa tgtgcactgc atagacacca ctgtatgcaa ttacagaagc 60
ctgtgagtga acggggtggg ggtcagtgcg ggcccatggc ctggctgtgc atttacggaa 120
30 gtctatcagt gaattggggt gtggtcagtg cgggcccattg gcctggctgg gcctgggagg 180
ttctctgatg tgtgaggcag gaggggaagg agggtagggg atagacagtg ggagccccc 240
ccctggaaga cataacagta agtccaggcc cgaaggggcag caggatgctt gggggcccag 300
cttgggcggc ggggatgatg gagggcctgg ccagggtggc agggatgatg ggggcccag 360
ctggggtggc aggggtgatg gggggggctg gtctgggtgg cggggaagat ggggaagcct 420
35 ggctgggccc cctcctcccc tgccctccac ctgcagccgt ggatccggat gtgcttccct 480
gggtgcacatc ctctgggcca tcagctttca tggagggtgg gggcaggggc atgacaccat 540
cctgtataaa atccaggatt cctcctcctg aacgcccaca ctcagggtga aagtcacatt 600
ccgctcttgg ccattctctt aagagtagac caggattctg atctctgaag ggtgggtagg 660
gtggggcagt ggggggtgtg gacacaggag gcttcagggt ggggctgggt atgtctcttc 720
40 atcctcttat catctcccag tctcatctct catcctctta tcatctccca gtctcatctg 780
tcttctctct atctccagc ctcatctgtc atcctcttac catctcccag tctcatctct 840
tatcctctta tctcctagtc tcatccagac ttacctccca gggcgggtgc caggctcgca 900
gtggagcttg acatacgtcc ttctcaggc agaaggaaat ggaaggattg cagagaacag 960
gagggggcgc tcagagggac gcagctcttg ggtgaagaaa cagccctccc tcagaagtgt 1020
45 gcttgggcca cacgaaaccg agggccctgc gtgagtggct ccagagcctt ccagcaggct 1080
cctgggtggg ccttatggta tggccgggtc ctactgagtg caccctggac agggcttctg 1140
gtttgagtgc agcccggacg tgcctgggtg cggggtgggg gcttatggcc actggatatg 1200
gcgtcattta ttgctgctgc ttcagagaat gtctgagtga ccgagcctaa tgtgtatgg 1260
gggcccagat ccaagactg tgcgtaaat gcactctgtt gcctggagcc cccgtatagg 1320
50 agctgtgagg aaggaggggc tcttggcagc cggcctgggg gcgcctttgc cctgcaaat 1380
ggaagggagc gggcccgggc gccgtgggag gacgacctca agtgagaggg tggacagaa 1440
agggcgggga cttcccagga gcagaggccg ctgctcaggc acacctgggt ttgaatcaca 1500
gaccaacagg tcaggccatt gttcagctat ccatcttcta caaagctcca gattcctgtt 1560
tctccgggtg tttttgttg aaattttact caggattact tatatttttt gctaaagtat 1620
55 tagacctta aaaaaggtat ttgctttgat atggcttaac tactaagca cctactttat 1680
ttgtctgttt ttatttatta ttattattat tattagagat ggtgtctact ctgtcaccca 1740
ggttgttagt gcagtggcac agtcatggct cgtgtagcc gcaaaccctc aggtcgaagt 1800
gatcctccgg cctcagcttc ccagagtgtt gggattacag gtgtgagcca ctgcccttgc 1860
ctggcacttt taaaaaccac tatgtaaggc cagggtccagt ggcttccaca cctgtcatcc 1920
60 cagtgtttt ggaagccgag gcagaaggat tgtctgagtc caggagtgtt agaccagcat 1980
gggtaacata gggagacccc atctctacaa aaaaagcaaa aagttatccg ggcgtgggg 2040
ccagcatctg tagtcccagc tgctcgggag gctgagtggg aggtatcgctt gagcccgga 2100
gggtcatggc gcagttagct gtgattgtac catcgactc cagcctgggc aacagagtga 2160
gacctgtct caaaaaaaaa aaaaaaaaaa gaaggagaag gagaagagaa gaagaaggaa 2220
65 gaaggaaaga gaagaagaag gaagaaggaa gaaagaaggaa gaaggaggcc tgctagggtc 2280

1
taggtagact gtcaaatctc agagcaaat gaaaataaca aagttttaa gggaaagaaa 2340
aaccacagct ctttggactt ccttaggcct gaacttcac tcaagcagct tccttccaca 2400
gacaagcgtg tatggagcga gtgagttcaa agcagaaagg gaggagaagc aggcaagggt 2460
5 ggaggctgtg ggtgacacca gccaggaccc ctgaaaggga gtggttgttt tcctgcctca 2520
gccccacgct cctgccggtc ctgcacctgc tctaaccgtc gatgttggtg ccagggtgcc 2580
acctgggaag gatgctgtgc agggggcttg ccaaactttg gtgggtttca gaagccccag 2640
gcacttgttg caggcacaat tacagcccct ccccaaagat gccacgtcc ttctcctgga 2700
acctgtgaat gtgtcaccgc caaggcagag gctggtgaag gctgcaggtg gaatcacggc 2760
10 tgccagtcag ccgatcttaa ggtcatcctg gattatctgg tgggcctgat atggccacaa 2820
gggtccctag aagtgaagaga gggaggcagg ggagagtcag agaggggacg tgagaaggac 2880
cactggccac tgctggcttt gagatggagg aggggggtccc cagccaagga atgggggcag 2940
ccgctccatg ctggaaaagc aagcaatcct ccccggtcct gagggcacac ggccctgccc 3000
acgcctcgat ttcaggccag tgggacctgt ttcagctttc cggcctccag agctgtaaga 3060
15 tgatgcgttt gtgttcagcc actaagctgc agtgattcgt cacagcagca aatggaatag 3120
cagtacaggg aaatgaatac agggacagtt ctacagtgta ctctcagccc acccctggg 3179